

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.52905 Seconds
(without alignment)

1830.070 Million cell updates/sec

Title: US-09-806-703A-14

Perfect score: 112

Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 segs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	112	100.0	21	AA11896	Immunogenic conjug
2	112	100.0	21	AA11896	Tetanus toxoid pro
3	112	100.0	21	AA11896	T-cell antigen TT3
4	112	100.0	21	AA11896	Broad range helper
5	112	100.0	21	AA11896	Tetanus toxin frag
6	112	100.0	21	AA11896	T-cell epitope pep
7	112	100.0	21	AA11896	Tetanus toxoid epi
8	112	100.0	21	AA11896	Tetanus P30 epitop
9	112	100.0	21	AA11896	Tetanus toxin T ce
10	112	100.0	21	AA11896	Foreign epitope P3

11	112	100.0	21	AA11896	Immunogenic conjug
12	112	100.0	21	AA11896	Tetanus toxoid pro
13	112	100.0	21	AA11896	T-cell antigen TT3
14	112	100.0	21	AA11896	Broad range helper
15	112	100.0	21	AA11896	Tetanus toxin frag
16	112	100.0	21	AA11896	T-cell epitope pep
17	112	100.0	21	AA11896	Tetanus toxoid epi
18	112	100.0	21	AA11896	Tetanus P30 epitop
19	112	100.0	21	AA11896	Tetanus toxin T ce
20	112	100.0	21	AA11896	Foreign epitope P3
21	112	100.0	21	AA11896	Immunogenic conjug
22	112	100.0	21	AA11896	Tetanus toxoid pro
23	112	100.0	21	AA11896	T-cell antigen TT3
24	112	100.0	21	AA11896	Broad range helper
25	112	100.0	21	AA11896	Tetanus toxin frag
26	112	100.0	21	AA11896	T-cell epitope pep
27	112	100.0	21	AA11896	Tetanus toxoid epi
28	112	100.0	21	AA11896	Tetanus P30 epitop
29	112	100.0	21	AA11896	Tetanus toxin T ce
30	112	100.0	21	AA11896	Foreign epitope P3
31	112	100.0	21	AA11896	Immunogenic conjug
32	112	100.0	21	AA11896	Tetanus toxoid pro
33	112	100.0	21	AA11896	T-cell antigen TT3
34	112	100.0	21	AA11896	Broad range helper
35	112	100.0	21	AA11896	Tetanus toxin frag
36	112	100.0	21	AA11896	T-cell epitope pep
37	112	100.0	21	AA11896	Tetanus toxoid epi
38	112	100.0	21	AA11896	Tetanus P30 epitop
39	112	100.0	21	AA11896	Tetanus toxin T ce
40	112	100.0	21	AA11896	Foreign epitope P3
41	112	100.0	21	AA11896	Immunogenic conjug
42	112	100.0	21	AA11896	Tetanus toxoid pro
43	112	100.0	21	AA11896	T-cell antigen TT3
44	112	100.0	21	AA11896	Broad range helper
45	112	100.0	21	AA11896	Tetanus toxin frag

ALIGNMENTS

RESULT 1
AA11896 standard; peptide; 21 AA.
AA11896;
19-JUL-1991 (first entry)
Immunogenic conjugate constituent peptide, TT3.
Malaria vaccine; major histocompatibility complex.
Tetanus toxin.
Key
Peptide
Location/Qualifiers
1..14
/label= active fragment (claimed)
BP427347-A.
15-MAY-1991.
07-NOV-1990; 90EP-0202948.
10-NOV-1989; 89IT-0022355.
(ENIE) ENRICECHER SPA.
Bianchi E, Pessi A, Corradin G;
WPI; 1991-141874/20.
Synthetic peptide(s) used as universal carriers - for preparing

Amino acid sequenc
CDA+ T cell epitop
Clostridium tetani
Amino acid sequenc
Universal tetanus
Tetanus Toxoid uni
Tetanus toxin T-ce
HER-2 B cell pepti
Tetanus toxoid TT3
Tetanus toxoid TT
Tetanus toxoid pre
Tetanus toxoid 947
PSMpep010 - P30 in
PSMpep011 - P30 in
PSMpep012 - P30 in
LHRH-containing im
Amyloid beta/tetan
Synthetic immunoge
Peptide comprising
A peptide which ma
Synthetic immunoge
Tetanus toxoid 830
Amyloid beta/tetan
Tetanus toxoid epi
Amyloid beta/tetan
Synthetic immunoge
Tetanus toxoid epi
Amyloid beta/tetan
Peptide comprising
A peptide which ma
Immunogenic branch
Immunogenic branch
Immunogenic branch
Tetanus toxoid epi

PT immunogenic conjugates used as vaccines against Plasmodium
 PT falciparum
 XX
 PS Claim 1, page 13, 16pp; English.
 XX
 CC This peptide corresponds to residues 947-967 of Tetanus toxin. It can
 CC be used as a universal carrier for the prepn. of an immunogenic
 CC conjugate. It is covalently bound to a peptide or polysaccharide
 CC hapten derived from a pathogen. This conjugate can be used as
 CC a vaccine for malaria. This peptide is recognised by different T-
 CC helper cell clones in association with alleles of the human MHC.
 CC It contains 2 epitopes: (a) 953-967, recognised by DR5-restricted
 CC clones; and (b) 947-960, recognised by all other DR and DP-
 CC restricted clones.
 XX
 SQ Sequence 21 AA;
 Query Match 100.0%; Score 112; DB 12; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FNNFTVSEFWLRVPKVSASHLE 21
 DB 1 FNNFTVSEFWLRVPKVSASHLE 21
 RESULT 2
 ID AAW06130 standard; Peptide; 21 AA.
 XX
 AC AAW06130;
 XX
 DT 07-FEB-1997 (first entry)
 XX
 DE Tetanus toxoid protein T-cell epitope.
 XX
 KW Cholesteryl ester transfer protein; CERP; antigen; vaccine;
 KM cardiovascular disease; atherosclerosis; tetanus toxoid;
 XX T-cell epitope.
 OS Clostridium tetani.
 XX
 PN WO9634888-A1.
 XX
 PD 07-NOV-1996.
 XX
 PF 01-MAY-1996; 96WO-US06147.
 XX
 PR 01-MAY-1995; 95US-0432483.
 XX
 PA (TCEL-) T CELL SCI INC.
 XX
 PI Rittershaus CM, Thomas LJ;
 DR WPI; 1996-506103/50.
 XX
 PT Cholesteryl ester transfer protein B cell epitope linked to T cell
 PT epitope - used to generate vaccine to regulate CERP activity for
 PT decreasing the risk of developing a cardiovascular disease e.g.
 PT atherosclerosis
 XX
 PS Claim 11, Page 43, 72pp; English.
 XX
 CC A helper T-cell epitope (AAW06130) comprises amino acids 947-967
 CC of tetanus toxoid protein. It can be utilised in novel peptide
 CC vaccines (see also AAW06129, AAW06132) also including B-cell
 CC epitope(s) from human or rabbit cholesteryl ester transfer
 CC protein (CERP) to elicit an immune response against endogenous
 CC CERP activity, thereby treating or preventing a cardiovascular
 CC disease, such as atherosclerosis.
 XX
 SQ Sequence 21 AA;

Query Match 100.0%; Score 112; DB 17; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FNNFTVSEFWLRVPKVSASHLE 21
 DB 1 FNNFTVSEFWLRVPKVSASHLE 21
 RESULT 3
 ID AAR88397 standard; Peptide; 21 AA.
 XX
 AC AAR88397;
 XX
 DT 12-JUN-1996 (first entry)
 XX
 DE T-cell antigen TT3 peptide.
 XX
 KW T-antigen; vaccine; antibody; T-cell; T-lymphocyte;
 KM alpha-helix; coiled-coil heterodimer; core peptide; subunit.
 XX
 OS Synthetic.
 XX
 PN WO9531480-A1.
 XX
 PD 23-NOV-1995.
 XX
 PF 18-MAY-1995; 95WO-CA00293.
 XX
 PR 18-MAY-1994; 94US-0245507.
 XX
 PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.
 XX
 PI Cachia PJ, Hodges RS, Houston ME, Irvin RT, Kay CM;
 PI Zhou NE;
 XX
 DR WPI; 1996-010880/01.
 XX
 PT Hetero-dimeric polypeptide immunogen in coiled-coil configuration
 PT with different antigens on each sub:unit - useful in vaccines and
 PT for antibody prodn.
 XX
 PS Claim 7, Page 62; 95pp; English.
 XX
 CC This T-cell antigen TT3 peptide may be attached to a core peptide
 CC contained in one of the 2 subunits of an alpha-helical coiled-coil
 CC heterodimer. Each core peptide is comprised of terminal and
 CC internal AA repeat sequences. This peptide antigen is attached
 CC to the core peptide through covalent linkages to certain AA of the
 CC internal repeats. The 2 subunits of the heterodimer are arranged
 CC in a stable alpha-helical coiled-coil configuration having a 1:1
 CC stoichiometry, and the peptide antigen is disposed toward the outer
 CC surfaces of the configuration. The heterodimer may be used as a
 CC synthetic vaccine (optionally multivalent) or to generate
 CC antibodies.
 XX
 SQ Sequence 21 AA;
 Query Match 100.0%; Score 112; DB 17; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FNNFTVSEFWLRVPKVSASHLE 21
 DB 1 FNNFTVSEFWLRVPKVSASHLE 21
 RESULT 4
 ID AAW46449 standard; Peptide; 21 AA.
 XX
 AC AAW46449;

XX	18-MAY-1998	(first entry)
DT		
XX	Broad range helper T cell epitope from the tetanus toxoid protein.	
DE		
XX	Cholesteryl ester transfer protein; CERP; cholesteryl ester;	
KM	high density lipoprotein; HDL; very low density lipoprotein; VLDL;	
KW	low density lipoprotein; LDL; T cell epitope; antibody;	
KM	DNA plasmid-based vaccine; broad range helper T cell epitope;	
KM	treatment; cardiovascular disease.	
XX		
OS	Clostridium tetani.	
XX		
PN	MO9741227-A1.	
XX		
PD	06-NOV-1997.	
XX		
PF	01-MAY-1997; 97WO-US07294.	
XX		
PR	21-FEB-1997; 97US-0802967.	
PR	01-MAY-1996; 96US-0640713.	
XX		
PA	(TCEL-) T CELL SCI INC.	
XX		
P1	Thomas LJ;	
XX		
DR	WPI: 1997-549731/50.	
XX		
PT	DNA plasmid-based vaccine encodes CERP B cell and helper T cell	
PT	epitope(s) - used for elevating high density lipoprotein levels, and	
PT	for treating cardiovascular disease	
XX		
PS	Disclosure; Page 44; 67bp; English.	
XX		
CC	The present sequence represents a broad range helper T cell epitope	
CC	of the tetanus protein. It can be used in DNA plasmid-based vaccines	
CC	against cholesteryl ester transfer proteins (CERPs). CERPs mediate the	
CC	transfer of cholesteryl esters from high density lipoprotein (HDL)	
CC	to very low density lipoprotein (VLDL) and low density lipoprotein (LDL),	
CC	and vice versa. An increased CERP activity produces an atherogenic	
CC	lipoprotein profile and induces atherosclerosis. A DNA plasmid-based	
CC	vaccine comprises sequences encoding at least one B cell epitope of CERP	
CC	linked in frame with at least one segment encoding a broad range helper	
CC	T cell epitope. The vaccines can be used to elevate the ratio of	
CC	circulating HDL to circulating LDL, VLDL or total cholesterol in a human.	
CC	It can also be used for decreasing the level of endogenous CERP activity	
CC	in a human. The vaccine can be used to produce anti-CERP antibodies in	
CC	vivo and for treating cardiovascular disease.	
XX		
XX		
SO	Sequence 21 AA;	
	Query Match 100.0%; Score 112; DB 18; Length 21;	
	Best Local Similarity 100.0%; Pred. No. 3.9e-12;	
	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
OY	1 FNNFTVSFWLRLVPKVSASHLE 21	
DB	1 FNNFTVSFWLRLVPKVSASHLE 21	
RESULT 5		
ID	AAW67034 standard; peptide; 21 AA.	
AAW67034		
AC	AAW67034;	
XX		
DT	15-DEC-1998 (first entry)	
XX		
DE	Tetanus toxin fragment (residues 947-967).	
XX		
KM	Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;	
KM	dendritic poly-Lysine; epitope; tumour.	
XX		

OS	Clostridium tetani.
XX	
PN	W09843677-A1.
PD	08-OCT-1998.
XX	
PF	27-MAR-1998; 98WO-EP01922.
XX	
PR	27-MAR-1997; 97US-0041726.
PA	(INSP) INST PASTEUR.
PI	Bay S., Cantacuzene D, Leclerc C, Lo-man R;
DR	WPI; 1998-557071/47.
XX	
PT	Carbohydrate peptide conjugate used as vaccine - comprises carrier
PT	with dendritic poly-Lysine enabling multiple epitopes to be
PT	covalently attached
XX	
XX	
XX	Disclosure; Page 13; 55pp; English.
PS	
CC	The invention relates to a new carbohydrate peptide conjugate, which
CC	comprises a carrier with a dendrimeric poly-lysine enabling multiple
CC	epitopes to be covalently attached to it. Also claimed are: (1) an
CC	antibody purified from biological fluid or cells of organisms
CC	administered with the carbohydrate peptide conjugate, and (2) a diagnosis
CC	kit comprising antigen-specific antibodies elicited by immunisation with
CC	the carbohydrate peptide conjugate. The peptide conjugate, antibody and
CC	diagnosis kit are used to provide pharmaceutical compositions and
CC	vaccines against tumours. These can be used to support an immune response
CC	against viral infections caused by hepatitis virus, HIV or cytomegalo
CC	virus. They can be used to enhance immune responses, especially B- and T-
CC	cell responses, of humans and animals against bacterial infections. The
CC	carbohydrate peptide conjugate stimulates the antibody and T-cell
CC	response without stimulating undesired immune responses. The composition
CC	is capable of increasing the survival of tumour bearing humans and
CC	animals. The present sequence corresponds to residues 947-967 of tetanus
CC	toxin. The synthetic peptide corresponding to this sequence may be used
CC	as an epitope in a carbohydrate peptide conjugate.
CC	
SQ	Sequence 21 AA:
	Query Match. 100.0%; Score 112; DB 19; Length 21;
	Best Local Similarity 100.0%; Pred. No. 3,9e-12;
	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY	1 FNNFTVSFWLRVPKVSASHLE 21 1 FNNFTVSFWLRVPKVSASHLE 21
DB	
RESULT 6	
ID	AAW67579 standard; peptide; 21 AA.
XX	AAW67579;
AC	
XX	
DT	02-MAR-1999 (first entry)
DE	T-cell epitope peptide #5 for chimeric fimbriin/T-cell epitope peptide.
XX	
KX	Chimeric; non-Typable Haemophilus influenzae; fimbriin; T-cell epitope;
KW	immunogenic composition; immune response.
OS	Synthetic.
XX	
PN	US5843464-A.
XX	
PD	01-DEC-1998.
XX	
PF	02-JUN-1995; 95US-0460502.
XX	

PR 02-JUN-1995; 95US-0460502.
 XX (OHIS) UNIV OHIO STATE.
 PA Bakaletz LO, Kaumaya PTP;
 PI WPI; 1999-044514/04.
 DR WPI; 1999-044514/04.
 XX Synthetic chimeric fimbria peptide - useful for vaccination against
 PT non-typable Haemophilus influenzae
 XX Disclousure; Column 4; 16pp; English.
 PS The invention relates to the manufacture of a synthetic chimeric peptide
 CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via
 CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is
 CC used in immunogenic compositions which induce an immune response against
 CC non-typable Haemophilus influenzae. This sequence represents an example
 CC of a T-cell epitope peptide used to generate the chimeric peptide.
 XX Sequence 21 AA:

Query Match 100.0%; Score 112; DB 20; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3,9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNNFTVSFWMRLRVPKVSASHLE 21
 DB 1 FNNFTVSFWMRLRVPKVSASHLE 21

RESULT 7

AAW73222
 ID AAW73222 standard; Protein; 21 AA.

XX AAW73222;

DT 25-JAN-1999 (first entry)

XX Tetanus toxoid epitope.

XX Multispecific single chain antibody; antibody H22; tumour cell; therapy;
 KM antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
 KM epidermal growth factor receptor; breast cancer; ovarian cancer.

XX Synthetic.

XX US5837243-A.

PD 17-NOV-1998.

PF 07-JUN-1996; 96US-0661052.

PR 07-JUN-1996; 96US-0661052.

PR 07-JUN-1995; 95US-0484172.

XX (MEDA-) MEDAREX INC.

PI Deo YM, Goldstein J, Graziano R, Somasundaram C;

DR WPI; 1999-023374/02.

XX Specific killing of tumour cells - using a multi-specific molecule
 PT comprising an anti-Fc receptor antibody and a portion which binds to
 PT a target cell

XX Example 7; Column 27; 57pp; English.

CC This sequence represents a tetanus toxoid epitope and is recognised
 CC by the multispecific single chain antibody designated H22. The
 CC antibody can be used in the method of the invention for inducing
 CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
 CC which is characterised by overexpression of HER 2/neu or epidermal growth

CC factor receptor (EGFR), comprises contacting the tumour cell with a
 CC multispecific protein molecule (preferably a single chain antibody)
 CC comprising: (a) an anti-Fc receptor antibody or an antigen binding
 CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
 CC binds to EGFR. The method can be used for treating cancers especially
 CC breast cancer or ovarian cancer. The multispecific antibody can also
 CC be administered prophylactically to vaccinate a subject against infection
 CC by a target cell.

XX Sequence 21 AA:

Query Match 100.0%; Score 112; DB 20; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3,9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNNFTVSFWMRLRVPKVSASHLE 21
 DB 1 FNNFTVSFWMRLRVPKVSASHLE 21

RESULT 8

AAB45512
 ID AAB45512 standard; Protein; 21 AA.

XX AAB45512;

DT 26-FEB-2001 (first entry)

XX Tetanus p30 epitope SEQ ID NO: 24.

XX Asthma; IL-5; interleukin-5; allergy; cytokine; helminthic infection;
 KM cancer; eosinophilia; vaccine; allergic rhinitis.

XX Clostridium tetani.

XX WO200065058-A1.

PD 02-NOV-2000.

PF 19-APR-2000; 2000WO-DK00205.

PR 23-APR-1999; 99DK-0000552.

PR 06-MAY-1999; 99US-0132811.

XX (MEBT-) M & E BIOTECH AS.

XX Klynsner S;

DR WPI; 2000-672791/65.

PT Down-regulating interleukin 5 (IL-5) activity in humans by
 PT administering IL-5 and/or an IL-5 analogue, useful in the treatment,
 PT prophylaxis or amelioration of asthma or other chronic allergic
 PT conditions -

XX Example 6; Page 137; 172pp; English.

CC The present invention is concerned with methods of treating asthma,
 CC eosinophilia, allergic rhinitis and other allergic diseases. These
 CC involve the use of interleukin-5 (IL-5) analogues and modified IL-5
 CC proteins and their coding sequences to down-regulate IL-5 activity and
 CC thus reduce eosinophil numbers. The allergic diseases may be treated
 CC using autovaccines, nucleic acid vaccines or live vaccines. In addition,
 CC it is possible that they may be used in the treatment of cancer and
 CC helminthic infections.

XX Sequence 21 AA:

Query Match 100.0%; Score 112; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3,9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNNFTVSFWMRLRVPKVSASHLE 21

Db 1 FNNFTVSFWLVRPKVSASHLE 21

RESULT 9
AAV99876

ID AAV99876 standard; Protein; 21 AA.

XX AAV99876;

DT 03-OCT-2000 (first entry)

DE Tetanus toxin T cell epitope helper peptide P30.

XX Human; MAGE-10; tumour rejection antigen precursor; bladder cancer;
KW prostate cancer; lung cancer; cancer detection; oesophageal cancer;
KM head and neck cancer; melanoma; myeloma; sarcoma; immunogen;
XX tetanus toxin.

XX Homo sapiens.

PN WO200026407-A1.

PD 11-MAY-2000.

PF 15-OCT-1999; 99WO-US24258.

PR 30-OCT-1998; 98US-0183714.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Boon-Falleur T, Brasseur F, Rimoldi D, Deplaen E;

DR WPI; 2000-451624/39.

PT Determining presence of cancer in samples, especially useful for
PT detecting bladder, prostate and lung cancer comprises assaying sample
PT for expression of tumor rejection antigen precursor MAGE-10 -

PS Example 12; Page 14; 26pp; English.

XX The present sequence is a tetanus toxin T cell epitope known as
CC Helper peptide P30. Hybrids of this peptide and an immunogenic peptide
CC derived from tumour rejection antigen precursor MAGE-10 were used to
CC generate polyclonal antiserum against MAGE-10. MAGE-10 binding monoclonal
CC antibodies can be used to detect MAGE-10 expression. A correlation
CC between MAGE-10 expression and cancer has been discovered and thus by
CC determining the presence of MAGE-10, the presence of cancer can be
CC determined. MAGE-10 expression can be detected using an immunoassay, an
CC oligonucleotide hybridisation assay or via other standard techniques.
CC This method is especially useful for determining the presence of
CC bladder, oesophageal, head and neck, prostate or lung cancer, or
CC melanoma, myeloma or sarcoma.

XX Sequence 21 AA;

Query Match 100.0%; Score 112; DB 21; Length 21;

Best Local Similarity 100.0%; Pred. No. 3.9e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLVRPKVSASHLE 21

DB 1 FNNFTVSFWLVRPKVSASHLE 21

RESULT 10
AAV92626

ID AAV92626 standard; Protein; 21 AA.

XX AAV92626;

DT 10-AUG-2000 (first entry)

DE Foreign epitope P30.

XX Foreign epitope; P2; prostate specific membrane antigen; PSM; Her2;
KW Heregulin 2; Fibroblast growth factor 8b; FGF8b; vaccination;
KW cytotoxic T-lymphocyte immunity; self-protein; cancer; breast cancer;
KM prostate cancer; cell-associated peptide antigen.

XX Clostridium tetani.

PN WO200020027-A2.

PD 13-APR-2000.

PF 05-OCT-1999; 99WO-DK00525.

PR 05-OCT-1998; 98DK-0001261.

PR 20-OCT-1998; 98US-0105011.

PA (MEBI-) M & B BIOTECH AS.

PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;

DR N-PSDB; AAA09461.

DR WPI; 2000-349917/30.

PT Inducing immune responses to weakly immunogenic, tumor associated

PT peptide antigens for the treatment of breast and prostate cancer

PS Example 1; Page 214; 220pp; English.

XX The claims detail a method for inducing immune responses against weakly
CC immunogenic cell-associated peptide antigens (PA) such as those
CC associated with cancers (i.e. self-proteins), for example, human
CC prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or
CC fibroblast growth factor 8b (FGF8b). The method comprises effecting
CC simultaneous presentation by antigen producing cells (APCs) of the
CC animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte)
CC group derived from the PA and/or at least 1 B-cell group derived from the
CC cell-associated PA; and (2) at least 1 first T helper cell group which is
CC foreign to the animal. Analogues of human PSM, human Her2 and
CC human/murine FGF8b comprising a substantial part of all known and
CC predicted CTL and B-cell epitopes of the respective PA and including at
CC least one foreign T helper epitope (e.g. P2 and/or P30) are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
XX when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 21 AA;

Query Match 100.0%; Score 112; DB 21; Length 21;

Best Local Similarity 100.0%; Pred. No. 3.9e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLVRPKVSASHLE 21

DB 1 FNNFTVSFWLVRPKVSASHLE 21

RESULT 11
AAV84428

ID AAV84428 standard; Peptide; 21 AA.

XX AAV84428;

DT 25-JUL-2000 (first entry)

DE Amino acid sequence of the tetanus toxoid P30 epitope.

KM Osteoprotegerin ligand; OPG; osteoprotegerin; osteoclastogenesis;

KM tumour necrosis factor receptor; type II transmembrane protein;

KM osteoclast differentiation; CSF-1; osteoclast activator;
KM immune response; osteoporosis; bone resorption;
KM tetanus toxoid P30 epitope.

```

XX OS Clostridium tetani.
XX PN WO200015807-A1.
XX PD 23-MAR-2000.
XX PF 13-SEP-1999; 99WO-DK00481.
XX PR 15-SEP-1998; 98DK-0001164.
XX PR 02-OCT-1998; 98US-0102896.
XX PA (MEBI-) M & E BIOTECH AS.
XX PI Halkier T, Haaning J;
XX DR WPI; 2000-271444/23.
XX PT In vivo down-regulation of osteoprotegerin ligand (OPGL) activity used
XX PS to treat, prevent and ameliorate osteoporosis -
XX PS Example: Page 106; 110pp; English.
XX CC The present sequence represents the tetanus toxoid P30 epitope. It is
XX CC used to create a fusion protein with murine osteoprotegerin ligand
XX CC (OPGL). Osteoprotegerin is a secreted member of the tumour necrosis
XX CC factor receptor family, which blocks osteoclastogenesis in a dose
XX CC dependent manner. The OPGL protein is synthesised as a type II
XX CC transmembrane protein. The murine and human OPGL polypeptides are 87%
XX CC homologous. OPGL is a potent osteoclast differentiation factor when
XX CC combined with CSF-1. It is not capable of inducing osteoclast
XX CC differentiation in the absence of CSF-1. OPGL is also an activator of
XX CC mature osteoclasts. The specification describes a method for the in vivo
XX CC down-regulation of OPGL activity in an animal. The method comprises
XX CC using at least one OPGL polypeptide or subsequence, and/or at least one
XX CC OPGL analogue to induce an immune response in the animal. The method
XX CC and OPGL polypeptide are useful for treating, preventing and ameliorating
XX CC osteoporosis or other diseases or conditions characterised by excessive
XX CC bone resorption.
XX SQ Sequence 21 AA;
XX
XX Query Match 100.0%; Score 112; DB 21; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-12;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 FNNFTVSFWLRVPKVSASHLE 21
XX Db 1 FNNFTVSFWLRVPKVSASHLE 21
XX
XX RESULT 12
XX ID AAY49260 standard; peptide; 21 AA.
XX AC AAY49260;
XX DT 07-FEB-2000 (first entry)
XX DE CD4+ T cell epitope P30TT fragment.
XX KW Carrier protein; CD4+; T cell epitope; N6; N10; N19; immunogen;
XX KW encapsulated bacteria.
XX OS Synthetic.
XX PN WO9955730-A2.
XX PD 04-NOV-1999.
XX PR 27-APR-1999; 99WO-1B00844.
XX PR 27-APR-1998; 98GB-0008932.

```

```

XX XX (CHTR-) CHIRON SPA.
XX PA Rappuoli R, Grandi G;
XX PI WPI; 2000-023325/02.
XX DR
XX PT Carrier proteins containing CD4+ epitopes useful for protecting against
XX PS diseases caused by encapsulated bacteria -
XX PS Disclosure; Page 36; 76pp; English.
XX CC The invention provides carrier proteins comprising at least 5 CD4+ T
XX CC cell epitope. The carrier protein comprises at least 1 of N6, N10 or
XX CC N19. The carrier protein can be prepared by expressing a vector
XX CC comprising a nucleic acid molecule encoding the protein, in a host cell
XX CC and recovering the expressed protein. The carrier protein can also be
XX CC produced by (a) constructing oligonucleotide molecules that encode
XX CC peptide epitopes; (b) annealing the oligonucleotides to form duplexes;
XX CC (c) introducing the duplexes into an expression vector; (d) introducing
XX CC the expression vector into a host cell; and (e) isolating the fusion
XX CC protein produced from a culture of the host cells. The carrier protein
XX CC can be used as a protective immunogen in the control of diseases caused
XX CC by encapsulated bacteria. Sequences AAY49256-266 represent CD4+ T cell
XX CC epitopes inserted in the recombinant polypeptide carrier proteins.
XX SQ Sequence 21 AA;
XX
XX Query Match 100.0%; Score 112; DB 21; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-12;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 FNNFTVSFWLRVPKVSASHLE 21
XX Db 1 FNNFTVSFWLRVPKVSASHLE 21
XX
XX RESULT 13
XX ID AAE11764 standard; peptide; 21 AA.
XX AC AAE11764;
XX DT 18-DEC-2001 (first entry)
XX DE Clostridium tetani P30 epitope.
XX KW Amyloid protein; neuroprotective; nootropic; immunostimulant; vaccine;
XX KW Alzheimer's disease; anticonvulsant; gene therapy; Pick's disease;
XX KW antidiabetic; systemic amyloidosis; maturity onset diabetes; ALS;
XX KW amyotrophic lateral sclerosis; Parkinson's disease; encephalopathy;
XX KW Huntington's disease; fronto-temporal dementia; P30 epitope.
XX OS Clostridium tetani.
XX PN WO200162284-A2.
XX PD 30-AUG-2001.
XX PF 19-FEB-2001; 2001WO-DK00113.
XX PR 21-FEB-2000; 2000DK-0000265.
XX PR 01-MAR-2000; 2000US-186295P.
XX PA (MEBI-) M & E BIOTECH AS.
XX PI Birk P, Jensen MR, Nielsen KG;
XX PR WPI; 2001-589796/66.
XX PR N-PSDB; AAD18756.
XX PT In vivo down-regulation of amyloid protein for the treatment of
XX PT Alzheimer's, comprises presenting an amyloidogenic polypeptide or its

```

PT subsequence and/or at least one analogue of the amyloidogenic
 PT polypeptide to the immune system -
 XX
 PS Example 3; Page 118; 120pp; English.
 CC The invention relates to a method for in vivo down-regulation of amyloid
 CC protein such as beta amyloid (Abeta) in an animal, including human. The
 CC method comprising presenting to the animal's immune system an
 CC immunogenically effective amount of at least one amyloidogenic protein
 CC or its subsequence and/or at least one analogue of the amyloidogenic
 CC polypeptide. The amyloidogenic protein or its subsequence, and its
 CC analogue is useful for the preparation of an immunogenic composition
 CC comprising an adjuvant for down-regulating amyloid in an animal. They are
 CC also useful in the treatment, prophylaxis or amelioration of Alzheimer's
 CC disease or other diseases characterised by amyloid deposits. They are
 CC also useful in the treatment of systemic amyloidosis, maturity onset
 CC diabetes, Parkinson's disease, Huntington's disease, fronto-temporal
 CC dementia, amyotrophic lateral sclerosis (ALS), Pick's disease and
 CC prion-related transmissible spongiform encephalopathies. They are also
 CC useful for inducing production of antibodies against an amyloidogenic
 CC polypeptide. The present sequence is Clostridium tetani p30 epitope
 CC related to the invention.
 CC
 SQ Sequence 21 AA;
 CC
 QY Query Match 100.0%; Score 112; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 FNNFTVSFWLRVPKVSASHLE 21
 DB 1 FNNFTVSFWLRVPKVSASHLE 21
 RESULT 14
 ID AAB85702 standard; peptide; 21 AA.
 AC AAB85702;
 XX
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of p30 epitope.
 XX
 KW Multivalent protein; immune response; Plasmodium vivax; parasite;
 KM protozoacide; vaccine; malaria; recombinant; ViVac1, ViVac2.
 XX
 OS Plasmodium vivax.
 PN WO20015181-A2.
 XX
 PD 02-AUG-2001.
 PF 29-JAN-2001; 2001WO-US02937.
 PR 31-JAN-2000; 2000US-0179213.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Lal AA, Xiao L, Zhou Z;
 PI WPI; 2001-514557/56.
 DR WPI; 2001-514557/56.
 XX
 PT New recombinant multivalent protein comprising antigenic determinants
 PT derived from more than one stage in a life cycle of Plasmodium vivax,
 PT useful as a vaccine for treating, preventing and reducing malaria
 PT infection -
 XX
 PS Example 1; Page 25; 59pp; English.
 CC The invention relates to recombinant multivalent proteins (I) that
 CC stimulate an immune response to Plasmodium vivax. (I) comprises antigenic
 CC determinants, fragments or conservative substitutions, derived from more

CC than one stage in a life cycle of a Plasmodium vivax parasite. (I) is
 CC useful as a vaccine for stimulating an immune response, specifically a
 CC protective immune response that confers increased resistance to infection
 CC by Plasmodium parasites, such as P. vivax. (I) is especially useful in
 CC the treatment, prevention and reduction of malarial infection, as
 CC research or diagnostic reagents for the detection of Plasmodium species
 CC in a biological sample, and for conferring immunity against multiple
 CC stages of the malarial parasite. The antibodies produced are useful for
 CC the detection or measurement of antigenic epitopes derived from one or
 CC more stages in a life cycle of a parasite, particularly P. vivax. The
 CC vaccine comprising the recombinant proteins, is cost-effective, health-
 CC promoting intervention for controlling, preventing or treating the
 CC incidence of malaria. The present sequence represents the amino acid
 CC sequence of a p30 epitope, a component of the multivalent and
 CC multistage proteins ViVacip and ViVac2p.
 CC
 SQ Sequence 21 AA;
 CC
 QY Query Match 100.0%; Score 112; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3.9e-12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 FNNFTVSFWLRVPKVSASHLE 21
 OY 1 FNNFTVSFWLRVPKVSASHLE 21
 RESULT 15
 ID AAB85453 standard; peptide; 21 AA.
 AC AAB85453;
 XX
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE Universal tetanus toxin Th epitope T9947-967.
 XX
 KW HER 2/neu; epidermal growth factor receptor; EGFR; multispecific protein;
 KM Fc receptor; FcR; tumor cell; breast; cancer; sarcoma; carcinoma; HIV;
 KM pathogenic; Toxoplasma gondii; candidiasis; systemic lupus; cytostatic;
 KM immune thrombocytopenia purpura; immunosuppressive; antiviral;
 KM antifungal; antiprotosomal; tetanus toxin.
 XX
 OS Clostridium tetani.
 PN US6270765-B1.
 XX
 PD 07-AUG-2001.
 PF 06-NOV-1998; 98US-0188082.
 PR 07-JUN-1996; 96US-0661052.
 PR 07-JUN-1995; 95US-0484172.
 PA (MEDA-) MEDAREX INC.
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
 PI WPI; 2001-475189/51.
 DR WPI; 2001-475189/51.
 XX
 PT Inducing killing of tumor cells which expresses HER 2/neu or epidermal
 PT growth factor receptor (EGFR) by contacting the cell with multispecific
 PT proteins comprising an anti-Fc receptor, -Her 2/neu or -EGFR antibody,
 PT useful for treating cancer -
 XX
 PS Example 7; Column 29; 57pp; English.
 CC The invention relates to a new method for inducing killing of a tumor
 CC cell which expresses HER 2/neu or epidermal growth factor receptor
 CC (EGFR). The method comprises contacting the tumor cell with a
 CC multispecific protein comprising a component, preferably an antibody,
 CC which binds to an Fc receptor (FcR), Her 2/neu or EGFR. The method is
 CC useful for inducing killing of a tumor cell from breast cancer, sarcoma,

CC carcinoma, or ovarian cancer. Specific multispecific proteins can also be
CC administered to a subject to treat or prevent other diseases or
CC conditions, including pathogenic infections (e.g., viral (such as HIV),
CC protozoan infections (such as Toxoplasma gondii), fungal infections
CC (such as candidiasis), and an autoimmunity (e.g., immune thrombocytopenia
CC purpura and systemic lupus). The present sequence represents an universal
CC tetanus toxin Th epitope TT947-967.

XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 112; DB 22; Length 21;

Best Local Similarity 100.0%; Pred. No. 3.9e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRLVPKVSASHLE 21

Db 1 FNNFTVSFWLRLVPKVSASHLE 21

Search completed: July 22, 2003, 08:08:16
Job time : 2.52905 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.09218 Seconds
(without alignments)
1830.070 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

A_Geneseq.101002.*
1: /SID52/gcgdata/geneseq/emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	74	100.0	15 11 AAR06310	Tetanus toxin epit
2	74	100.0	15 18 AAM35506	Universal T-cell e
3	74	100.0	15 18 AAM11505	Tetanus toxoid uni
4	74	100.0	15 19 AAM67033	Tetanus toxin frag
5	74	100.0	15 19 AAM71321	Universal helper T
6	74	100.0	15 20 AAY04051	T-helper epitope f
7	74	100.0	15 20 AAM67578	T-cell epitope pep
8	74	100.0	15 20 AAM73220	Tetanus toxoid epi
9	74	100.0	15 21 AAB45511	Tetanus P2 epitope
10	74	100.0	15 21 AAY82637	Tetanus toxoid T c

11	74	100.0	15 21 AAY92625	Foreign epitope P2
12	74	100.0	15 21 AAY64427	Amino acid sequenc
13	74	100.0	15 21 AAY70300	Clostridium tetani
14	74	100.0	15 21 AAY44763	Tetanus toxoid pro
15	74	100.0	15 22 AAF11763	Clostridium tetani
16	74	100.0	15 22 AAM99515	Vaccine related MH
17	74	100.0	15 22 AAB85701	Amino acid sequenc
18	74	100.0	15 22 AAB85451	Wld-type TT830 (t
19	74	100.0	15 22 AAB61956	Tetanus Toxoid uni
20	74	100.0	15 22 AAB20143	Tetanus toxin T-ce
21	74	100.0	15 22 AAB6836	HER-2 B cell pepti
22	74	100.0	15 22 AAB46172	Tetanus toxoid TT8
23	74	100.0	15 22 AAB49071	Tetanus toxoid TT
24	74	100.0	15 23 AAY97872	Tetanus toxin P2 (
25	74	100.0	16 18 AAM35445	T-cell stimulatory
26	74	100.0	16 20 AAY97905	Clostridium tetani
27	74	100.0	16 23 AAY93865	Clostridium tetani
28	74	100.0	16 23 AAY11413	Tetanus toxoid pre
29	74	100.0	17 23 AAB2692	Helper T cell epit
30	74	100.0	17 16 AAR82573	Tetanus toxin help
31	74	100.0	17 17 AAM05599	Tetanus toxin help
32	74	100.0	17 17 AAR88395	T-cell antigen TT2
33	74	100.0	17 21 AAY9274	HIV class II bindi
34	74	100.0	17 21 AAY80056	Pathogen derived T
35	74	100.0	17 21 AAY54539	T helper cell (Th)
36	74	100.0	17 21 AAY58768	Unidentified pepti
37	74	100.0	17 22 AAM99516	Vaccine related MH
38	74	100.0	17 22 AAG52904	Amino acid residue
39	74	100.0	17 22 AAB84435	Amino acid sequenc
40	74	100.0	17 22 AAB30941	Amino acid sequenc
41	74	100.0	17 22 AAB31029	Antigenic fragment
42	74	100.0	17 22 AAB31118	Peptide 5 for pept
43	74	100.0	17 22 AAB15589	HIV-derived lipope
44	74	100.0	18 20 AAY26607	Peptide TT functio
45	74	100.0	18 23 AAB09794	

ALIGNMENTS

RESULT 1
AAR06310 standard; protein; 15 AA.
ID AAR06310;
XX AAR06310;
XX
AC AAR06310;
XX
DT 04-DEC-1990 (first entry)
XX
XX Tetanus toxin epitope.
XX
DE Tetanus toxin epitope.
XX
KM Tetanus toxin; vaccine; major histocompatibility complex; MHC;
XX
XX antimetallat.
XX
OS Synthetic.
XX
XX BP378881-A.
XX
XX PD 25-JUL-1990.
XX
XX PF 27-DEC-1989; 89BP-0203318.
XX
XX PR 16-NOV-1989; 89IT-0022409.
XX
XX PR 17-JAN-1989; 89IT-0019110.
XX
XX PA (ENIE) ENRICERCH SPA.
XX
XX PI Pessi A, Bianchi E, Verdini AS, Corradin G;
XX
XX DR WPI; 1990-225582/30.
XX
XX PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
XX as universal carriers for prepn. of immunogenic conjugate(s) for
XX use as vaccines.

XX Claim 1; Page 17; 20pp; English.
PS
XX
CC Epitopic peptides may be used with synthetic hapten derived from
CC a pathogen to generate an immune response to the pathogen.
CC Peptides are recognised by numerous T-helper cell clones within
CC the context of a wide range of alleles of the human MHC.
CC The peptides may be used in an antimalarial vaccine inducing Ab.
CC response to P.falciparum.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 74; DB 11; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYKANSKFIGITEL 15
DB 1 QYKANSKFIGITEL 15
RESULT 2
AAW35506
ID AAW35506 standard; peptide; 15 AA.
XX
AC AAW35506;
XX
DT 22-APR-1998 (first entry)
XX
DE Universal T-cell epitope peptide SEQ ID NO:8.
XX
XX T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
KM scaffold; inhibition; metastasis; wound healing; solid phase.
XX
OS Unidentified.
XX
PN MO9738011-A1.
XX
PD 16-OCT-1997.
XX
PF 03-APR-1997; 97WO-DE00146.
XX
PR 03-APR-1996; 96DK-0000398.
XX
PA (PEPR-) PEPRSEARCH AS.
XX
PI Heegaard PMH, Jakobsen PH;
XX
DR WPI; 1997-512645/47.
XX
PT Non-dendritic peptide carrier linked to a solid phase - useful as a
PT diagnostic agent and as a scaffold for production of chemical
PT derivatives
XX
PS Example 20; Page 124; 262pp; English.
XX
CC A non-dendritic peptide carrier (A) has been developed which is coupled
CC through a linker to a solid phase, forming a complex of (A)-solid phase.
CC Where (A) comprises 10-50 amino acids capable of forming a secondary
CC structure in a benign buffer after liberation from the solid phase, and
CC further the (A)-solid phase complex comprises an immunogenic substance
CC and/or an immune mediator coupled on (A). The present sequence
CC represents a peptide used in an example from the present invention.
CC (A)-solid phase complex can be used as a scaffold for the production of
CC chemical derivatives, characterised by covalently attaching molecules at
CC attachment points. Alternatively (A) is used as a scaffold-peptide for
CC the incorporation into an Immunostimulating Complex (Iscom) resulting in
CC (A)-Iscom complex which is used for the chemical coupling of antigenic
CC substances in an aqueous solution by conjugation. (A) derivatised with
CC one or more peptides having fibronectin-, laminin- or vitronectin-like
CC binding activities can be used for the promotion of cell-attachment to
CC plastic surfaces, in particular to inhibit tumour growth and metastasis,
CC and for promotion of wound healing. Also a derivatised (A) can be used

CC for the selection of specifically-binding aptamers or as a diagnostic
CC agent. Such diagnostic-(A) molecules could be used to detect molecules
CC derived from or indicative of pregnancy or of a disease, such as an
CC infectious, autoimmune or cancerous disease.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 74; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYKANSKFIGITEL 15
DB 1 QYKANSKFIGITEL 15
RESULT 3
AAW11505
ID AAW11505 standard; Protein; 15 AA.
XX
AC AAW11505;
XX
DT 24-SEP-1997 (first entry)
XX
DE Tetanus toxoid universal Th epitope TT830.
XX
XX Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;
KM fusion protein; chimera; tetanus toxoid; helper T cell epitope;
KM antigen presentation; ds.
XX
OS Clostridium tetani.
XX
PN MO9640789-A1.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US09988.
XX
PR 07-JUN-1995; 95US-0484172.
XX
PA (MEDA-) MEDAREX INC.
XX
PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
XX
DR WPI; 1997-052242/05.
XX
DR N-PSDB; AAT58127.
XX
PT Recombinant, multi-specific anti-Fc receptor antibody molecules -
PT also comprise an anti-target portion, used for the treatment of
PT cancer, autoimmune disease and pathogenic infection
XX
PS Example 7; Fig 24; 115pp; English.
XX
CC Synthetic DNA coding for the wild-type universal Th epitope from
CC tetanus toxoid, designated TT830, was fused to the 3'-end of DNA
CC encoding heavy chain sequences from the humanised anti-Fc gamma RI
CC monoclonal antibody H22. The resulting fusion protein was shown to
CC be significantly more efficient in antigen presentation and T cell
CC stimulation than the TT830 epitope alone. A similar fusion
CC construct was prepared coding for a mutant, antagonistic form of the
CC epitope (designated TT833S) fused to the anti-Fc gamma RI. The
CC Fab22-TT833S is at least 100 times more effective than TT833S in
CC inhibiting T cell activation.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 74; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYKANSKFIGITEL 15
DB 1 QYKANSKFIGITEL 15


```

RESULT 4
AAW67033
ID AAW67033 standard; peptide; 15 AA.
XX
XX AAW67033;
AC
XX
XX 15-DEC-1998 (first entry)
DT
XX Tetanus toxin fragment (residues 830-844).
XX
XX Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
XX dendritic poly-lysine; epitope; tumour.
XX
XX Clostridium tetani.
XX
XX MO9843677-A1.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98MO-EP01922.
XX
XX 27-MAR-1997; 97US-0041726.
XX
XX (INSP ) INST PASTEUR.
XX
XX Bay S, Cantacuzene D, Leclerc C, Lo-man R;
XX WPI; 1998-557071/47.
XX
XX Carbohydrate peptide conjugate used as vaccine - comprises carrier
XX PT with dendritic poly-lysine enabling multiple epitopes to be
XX PT covalently attached
XX
XX PS Disclosure; Page 13; 55pp; English.
XX
XX The invention relates to a new carbohydrate peptide conjugate, which
XX CC comprises a carrier with a dendritic poly-lysine enabling multiple
XX CC epitopes to be covalently attached to it. Also claimed are: (1) an
XX CC antibody purified from biological fluid or cells of organisms
XX CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
XX CC kit comprising antigen-specific antibodies elicited by immunisation with
XX CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
XX CC diagnosis kit are used to provide pharmaceutical compositions and
XX CC vaccines against tumours. These can be used to support an immune response
XX CC against viral infections caused by hepatitis virus, HIV or cytomegalo
XX CC virus. They can be used to enhance immune responses, especially B- and T-
XX CC cell responses, of humans and animals against bacterial infections. The
XX CC carbohydrate peptide conjugate stimulates the antibody and T-cell
XX CC response without stimulating undesired immune responses. The composition
XX CC is capable of increasing the survival of tumour bearing humans and
XX CC animals. The present sequence corresponds to residues 830-844 of tetanus
XX CC toxin. The synthetic peptide corresponding to this sequence may be used
XX CC as an epitope in a carbohydrate peptide conjugate.
XX
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 74; DB 19; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 QYIKANSKFIGITEL 15
XX |||||
XX 1 QYIKANSKFIGITEL 15
XX
XX DB
XX
XX RESULT 5
XX AAW71321
XX ID AAW71321 standard; peptide; 15 AA.
XX
XX AC AAW71321;
XX
XX

```

```

DT 26-NOV-1998 (first entry)
XX
XX Universal helper T-cell epitope P2 derived from tetanus toxin.
XX
XX DE
XX Liver stage; Plasmodium; Navy Yoelli Liver Stage 3 antibody; NYLS3;
XX K/M hepatic and erythrocytic stage protein; PyHEP17; vaccine;
XX KM malaria parasite; tetanus toxin; P2; helper T-cell epitope.
XX
XX OS Synthetic.
XX OS Clostridium tetani.
XX
XX PN US814617-A.
XX
XX PD 29-SEP-1998.
XX
XX PF 07-OCT-1994; 94US-0319704.
XX
XX PR 07-OCT-1994; 94US-0319704.
XX
XX Y.. (USNA ) US SEC OF NAVY.
XX
XX PA
XX PI Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
XX DR WPI; 1998-541794/46.
XX
XX XX
XX Vaccine for protecting mammal against infection by malaria caused by
XX PT Plasmodium species - comprises a first nucleic acid encoding a first
XX PT polypeptide capable of eliciting an immune reaction against an
XX PT antigen expressed during the liver
XX
XX PS Disclosure; Column 12; 24pp; English.
XX
XX AAW71321-22 represent universal helper T-cell epitopes derived from
XX CC tetanus toxin. They are used to enhance host immune response to
XX CC vaccines. The specification describes a Plasmodium yoelli liver stage
XX CC 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
XX CC protein elicits a response from an Ig1 monoclonal antibody designated
XX CC Navy Yoelli Liver Stage 3 (NYLS3). This antibody does not recognise
XX CC sporozoites, but does recognise P. yoelli liver stage parasites. NYLS3
XX CC eliminates upto 90% of liver stage parasites. The specification
XX CC describes a vaccine for reducing the severity or incidence of infection
XX CC by a malaria parasite of the genus Plasmodium. The DNA vaccine comprises
XX CC exon 1 and part of exon 2 of the PyHEP17 gene.
XX
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 74; DB 19; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 QYIKANSKFIGITEL 15
XX |||||
XX 1 QYIKANSKFIGITEL 15
XX
XX DB
XX
XX RESULT 6
XX AAY04051
XX ID AAY04051 standard; peptide; 15 AA.
XX
XX AC AAY04051;
XX
XX 04-JAN-2000 (first entry)
DT
XX T-Helper epitope from tetanus toxoid.
XX
XX DE
XX Covalently reactive antigen analog; CRAA; catalytic antibody;
XX KM electrophilic reaction centre; phosphate; boronate; vaccine;
XX KM transition state analog; TSA; isostere; gp120; HIV-1; T-helper;
XX KM tetanus; toxoid; B-T-epitope.
XX
XX OS Clostridium tetani.
XX
XX PN MO9948925-A1.
XX

```

```

XX 30-SEP-1999.
PD 23-MAR-1999; 99WO-US06325.
XX 23-MAR-1999; 98US-0046373.
XX 23-MAR-1998; 98US-0046373.
XX (UYNE-) UNIV NEBRASKA.
XX Paul S, Gololobov G, Smith L;
XX WPI; 1999-591076/50.
XX New covalently reactive antigen analogs used for treating e.g.
PT autoimmune diseases, lymphoproliferative disorders, cancers, microbial
PT infections, ischemic and reperfusion injury or septic shock -
XX Disclosure; Page 86; 158pp; English.
XX The patent discloses new covalently reactive antigen analogs (CRAA)
CC of formula XI-Y-E-X2, in which XI and X2 represent peptide sequences
CC of an epitope of a disease-associated protein, Y is a positively
CC charged amino acid residue, preferably Lys or Arg, and E is an
CC electrophilic reaction centre, preferably a phosphonate or boronate
CC moiety. Depending on the identity of the epitope, the CRAA may be used
CC to stimulate production of catalytic antibodies specific for
CC predetermined antigens associated with particular medical disorders.
CC They may also be used to permanently inactivate endogenously produced
CC catalytic antibodies produced in certain autoimmune diseases as well as
CC in certain lymphoproliferative disorders.
CC Amongst the specifically exemplified CRAAs is one based on residues
CC 421-436 of a B-cell epitope of gp120 (see AAY04046) which may be used
CC to counter HIV-1 infections. When used as an immunogen, preferably this
CC CRAA is conjugated at its N-terminal to a T-helper epitope from tetanus
CC toxoid. The present sequence represents the T-helper epitope and
XX corresponds to residues 830-844 of the toxoid.
XX
XX Sequence 15 AA:
SO
Query Match 100.0%; Score 74; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 OYIKANSKFIGITEL 15
Db 1 OYIKANSKFIGITEL 15
RESULT 7
AAW67578
ID AAW67578 standard; peptide; 15 AA.
XX
AC AAW67578;
XX
DT 02-MAR-1999 (first entry)
XX
DE T-cell epitope peptide #4 for chimeric fimbria/T-cell epitope peptide.
XX
KM Chimeric; non-typable Haemophilus influenzae; fimbria; T-cell epitope;
KM immunogenic composition; immune response.
XX
OS Synthetic.
XX
PN US5843464-A.
XX
XX 01-DEC-1998.
XX
PF 02-JUN-1995; 95US-0460502.
XX
PR 02-JUN-1995; 95US-0460502.
XX
PA (OHIS ) UNIV OHIO STATE.
XX

```

```

PI Bakaletz LO, Kaumaya PTP;
XX
XX WPI; 1999-044514/04.
DR
XX Synthetic chimeric fimbria peptide - useful for vaccination against
PT non-typable Haemophilus influenzae
XX
XX Disclosure; Column 4; 16pp; English.
XX
XX The invention relates to the manufacture of a synthetic chimeric peptide
CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via
CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is
CC used in immunogenic compositions which induce an immune response against
CC non-typable Haemophilus influenzae. This sequence represents an example
CC of a T-cell epitope peptide used to generate the chimeric peptide.
XX
XX Sequence 15 AA:
SO
Query Match 100.0%; Score 74; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 OYIKANSKFIGITEL 15
Db 1 OYIKANSKFIGITEL 15
RESULT 8
AAW73220
ID AAW73220 standard; Protein; 15 AA.
XX
AC AAW73220;
XX
DT 25-JAN-1999 (first entry)
XX
DE Tetanus toxoid epitope.
XX
XX Multispecific single chain antibody; antibody H22; tumour cell; therapy;
KM antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
KM epidermal growth factor receptor; breast cancer; ovarian cancer.
XX
OS Synthetic.
XX
PN US5837243-A.
XX
PD 17-NOV-1998.
XX
PF 07-JUN-1996; 96US-0661052.
XX
XX 07-JUN-1996; 96US-0661052.
PR 07-JUN-1995; 95US-0484172.
XX
XX (MEDA-) MEDAREX INC.
XX
XX Deo YM, Goldstein J, Graziano R, Somasundaram C;
XX WPI; 1999-023374/02.
XX
XX Specific killing of tumour cells - using a multi-specific molecule
PT comprising an anti-Fc receptor antibody and a portion which binds to
PT a target cell
XX
XX Example 7; Column 27; 57pp; English.
XX
XX This sequence represents a tetanus toxoid epitope and is recognised
CC by the multispecific single chain antibody designated H22. The
CC antibody can be used in the method of the invention for inducing
CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
CC which is characterised by overexpression of HER 2/neu or epidermal growth
CC factor receptor (EGFR), comprises contacting the tumour cell with a
CC multispecific protein molecule (preferably a single chain antibody)
CC comprising: (a) an anti-Fc receptor antibody or an antigen binding
CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which

```

CC binds to EGFR. The method can be used for treating cancers especially
 CC breast cancer or ovarian cancer. The multispecific antibody can also
 CC be administered prophylactically to vaccinate a subject against infection
 CC by a target cell.

XX Sequence 15 AA;

Query Match 100.0%; Score 74; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
 DB 1 QYIKANSKFIGITEL 15

RESULT 9

AAB45511 standard; Protein; 15 AA.

XX AAB45511;

XX 26-FEB-2001 (first entry)

DE Tetanus P2 epitope SEQ ID NO: 23.

XX Asthma; IL-5; interleukin-5; allergy; cytokine; helminthic infection;
 KM cancer; eosinophilia; vaccine; allergic rhinitis.

XX Clostridium tetani.

XX MO200065058-A1.

XX 02-NOV-2000.

XX 19-APR-2000; 2000MO-DK00205.

XX 23-APR-1999; 99DK-0000552.

XX 06-MAY-1999; 99US-0132811.

XX (MEBI-) M & E BIOTECH AS.

XX Klynsner S;

XX WPI; 2000-672791/65.

XX Down-regulating interleukin 5 (IL-5) activity in humans by
 PT administering IL-5 and/or an IL-5 analogue, useful in the treatment,
 PT prophylaxis or amelioration of asthma or other chronic allergic
 PT conditions -

XX Example 1; Page 137; 172pp; English.

XX The present invention is concerned with methods of treating asthma,
 CC eosinophilia, allergic rhinitis and other allergic diseases. These
 CC involve the use of interleukin-5 (IL-5) analogues and modified IL-5
 CC proteins and their coding sequences to down-regulate IL-5 activity and
 CC thus reduce eosinophil numbers. The allergic diseases may be treated
 CC using autovaccines, nucleic acid vaccines or live vaccines. In addition,
 CC it is possible that they may be used in the treatment of cancer and
 CC helminthic infections.

XX Sequence 15 AA;

Query Match 100.0%; Score 74; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
 DB 1 QYIKANSKFIGITEL 15

RESULT 10
 AAY82637 standard; peptide; 15 AA.

XX AAY82637;

XX 07-AUG-2000 (first entry)

DE Tetanus toxoid T cell epitope peptide SEQ ID NO:13.

XX T cell epitope; B cell epitope; allergy; allergen; antigenic;
 KM antiallergic; antiasthmatic; antiinflammatory; dermatological;
 KM immunosuppressive; vaccine; rhinitis; sinusitis; bronchial asthma;
 KM atopic dermatitis; acute urticaria; chronic urticaria;
 KM gastro-intestinal syndrome; food allergen; oro-pharyngeal syndrome;
 KM anaphylactic reaction; drug hypersensitivity; allergic reaction.

XX Clostridium tetani.

XX Synthetic.

XX MO200006694-A2.

XX 10-FEB-2000.

XX 20-JUL-1999; 99WO-BE00092.

XX 30-JUL-1998; 98EP-0870167.

XX (UNIO) UCB SA.

XX Saint-Remy J, Jacquemin M;

XX WPI; 2000-422470/36.

XX New compound for prevention and treatment of allergies comprises at
 PT least one allergen antigenic determinant recognized by a B cell and at
 PT least one antigenic determinant which does not trigger T cell
 PT activation -

XX Example 6; Page 30; 50pp; English.

XX The present invention describes a compound (I) for the prevention and/or
 CC treatment of allergy. The compound comprises at least one allergen
 CC antigenic determinant (i) recognised by a B cell or an antibody secreted
 CC by a B cell of a non-atopic individual and at least one antigenic
 CC determinant (ii) different from the allergen that triggers T cell
 CC activation. (I) has antiallergic, antiasthmatic, antiinflammatory,
 CC dermatological and immunosuppressive activities, and can be used in a
 CC vaccine. (I) may be used in a pharmaceutical or cosmetic medicament to
 CC treat and/or prevent allergies or a disease of allergic origin,
 CC especially hypersensitivities. These include rhinitis, sinusitis,
 CC bronchial asthma, atopic dermatitis, some forms of acute and chronic
 CC urticaria, gastro-intestinal syndromes associated with the ingestion of
 CC food allergens, oro-pharyngeal syndrome, anaphylactic reactions
 CC associated with drug hypersensitivities and/or a mixture of these. The
 CC use of (I) in the treatment of allergic conditions avoids the need for
 CC drug treatment, which often causes undesirable side-effects. Also, prior
 CC art drug therapies alleviate symptoms, but do not influence their
 CC causes, however (I) actually combats the cause of an allergic reaction.
 CC The present sequence represents a peptide, which is used in an
 CC example from the present invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 74; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
 DB 1 QYIKANSKFIGITEL 15

```

RESULT 11
AA92625
ID AAY92625 standard; Protein; 15 AA.
XX
XX AAY92625;
XX
XX 10-AUG-2000 (first entry)
XX
XX Foreign epitope P2.
XX
XX Foreign epitope P2.
XX
XX Foreign epitope P2; prostate specific membrane antigen; PSM; Her2;
XX Heregulin 2; Fibroblast growth factor 8b; FGF8b; vaccination;
XX cytotoxic T-lymphocyte immunity; self-protein; cancer; breast cancer;
XX prostate cancer; cell-associated peptide antigen.
XX
XX Clostridium tetani.
XX
XX MO200020027-A2.
XX
XX 13-APR-2000.
XX
XX 05-OCT-1999; 99WO-DK00525.
XX
XX 05-OCT-1998; 98DK-0001261.
XX 20-OCT-1998; 98US-0105011.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Steinaa L, Mourielsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
XX Gautam A, Birk P, Karlsson G;
XX
XX WPI; 2000-349917/30.
XX N-PSDB; AAA09460.
XX
XX Inducing immune responses to weakly immunogenic, tumor associated
XX peptide antigens for the treatment of breast and prostate cancer
XX
XX Example 1; Page 213; 220pp; English.
XX
XX The claims detail a method for inducing immune responses against weakly
XX immunogenic cell-associated peptide antigens (PA) such as those
XX associated with cancers (i.e. self-proteins), for example, human
XX prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or
XX fibroblast growth factor 8b (FGF8b). The method comprises effecting
XX simultaneous presentation by antigen producing cells (APCs) of the
XX animal's immune system of: (i) at least 1 CTL (cytotoxic T-lymphocyte)
XX group derived from the PA and/or at least 1 B-cell group derived from the
XX cell-associated PA; and (2) at least 1 first T helper cell group which is
XX foreign to the animal. Analogues of human PSM, human Her2 and
XX human/murine FGF8b comprising a substantial part of all known and
XX predicted CTL and B-cell epitopes of the respective PA and including at
XX least one foreign T helper epitope (e.g. P2 and/or P3) are also claimed.
XX The method is used to treat prostate, prostate/breast or breast cancer
XX when the PA is human PSM, FGF8b and Her2, respectively.
XX
XX Sequence 15 AA;
SQ
Query Match 100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYKANSKFIGITEL 15
DB 1 QYKANSKFIGITEL 15

```

```

XX
XX Amino acid sequence of the tetanus toxoid P2 epitope.
XX
XX Osteoprotegerin ligand, OPG, osteoprotegerin; osteoclastogenesis;
XX tumour necrosis factor receptor; type II transmembrane protein;
XX osteoclast differentiation; CSF-1; osteoclast activator;
XX immune response; osteoporosis; bone resorption;
XX tetanus toxoid P2 epitope.
XX
XX Clostridium tetani.
XX
XX MO200015807-A1.
XX
XX 23-MAR-2000.
XX
XX 13-SEP-1999; 99WO-DK00481.
XX
XX 15-SEP-1998; 98DK-0001164.
XX 02-OCT-1998; 98US-0102896.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Halkier T, Haaning J;
XX
XX WPI; 2000-271444/23.
XX
XX In vivo down-regulation of osteoprotegerin ligand (OPG) activity used
XX to treat, prevent and ameliorate osteoporosis -
XX
XX Example; Page 106; 110pp; English.
XX
XX The present sequence represents the tetanus toxoid P2 epitope. It is
XX used to create a fusion protein with murine osteoprotegerin ligand
XX (OPG). Osteoprotegerin is a secreted member of the tumour necrosis
XX factor receptor family, which blocks osteoclastogenesis in a dose
XX dependent manner. The OPG protein is synthesised as a type II
XX transmembrane protein. The murine and human OPG polypeptides are 87%
XX homologous. OPG is a potent osteoclast differentiation factor when
XX combined with CSF-1. It is not capable of inducing osteoclast
XX differentiation in the absence of CSF-1. OPG is also an activator of
XX mature osteoclasts. The specification describes a method for the in vivo
XX down-regulation of OPG activity in an animal. The method comprises
XX using at least one OPG polypeptide or subsequence, and/or at least one
XX OPG analogue to induce an immune response in the animal. The method
XX CC and OPG polypeptide are useful for treating, preventing and ameliorating
XX osteoporosis or other diseases or conditions characterised by excessive
XX bone resorption.
XX
XX Sequence 15 AA;
SQ
Query Match 100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QYKANSKFIGITEL 15
DB 1 QYKANSKFIGITEL 15

```

```

RESULT 13
AA70300
ID AAY70300 standard; peptide; 15 AA.
XX
XX AAY70300;
XX
XX 06-JUN-2000 (first entry)
XX
XX Clostridium tetani tetanus toxoid T-cell epitope, P589.
XX
XX Recombinant protein; CDC/NIMH/VAC-1; multivalent; malaria; vaccine;
XX T-cell epitope; tetanus toxoid; antigenic epitope; treatment;
XX circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;
XX liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;
XX

```

KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;
 KM EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;
 KN Pf27; antiparasitic; prevention; anti-CDC/N1MALVAC-1 antibody.
 OS Clostridium tetani.
 XX
 XX
 PN WO200011179-A1.
 XX
 PD 02-MAR-2000.
 XX
 PP 19-AUG-1999; 99WO-US18869.
 XX
 PR 21-AUG-1998; 98US-0097703.
 XX
 PA (NAIM-) NAT INST IMMUNOLOGY.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Lal AA, Shi YP, Hasnain SE;
 XX
 DR WPI: 2000-237654/20.
 PT Novel recombinant protein as vaccine for treating malarial infection
 PT comprises antigenic peptides obtained from different stages of
 PT plasmidium falciparum life cycle -
 PS
 PS Claim 2; Page 17; 52pp; English.
 XX
 CC The present sequence is the tetanus toxoid T-cell epitope P589, derived
 CC from Clostridium tetani. It is used in the construction of recombinant
 CC protein CDC/N1MALVAC-1, which is a multivalent, multistage malarial
 CC vaccine. The recombinant protein comprises, melitin signal peptide,
 CC (His)6 tag, T-cell epitope from tetanus toxoid and 21 antigenic epitopes
 CC from circumsporozoite protein (CSP), sporozoite surface protein-2
 CC (SSP-2), liver stage antigen-1 (LSA-1), merozoite surface protein-1
 CC (MSP-1), MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding
 CC antigen-175 (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete
 CC specific antigen, Pf27. These epitopes were obtained at different stages
 CC of the life cycle of P. falciparum. CDC/N1MALVAC-1 vaccine has
 CC antiparasitic activity and can be used for treatment and prevention of
 CC malarial infections. Anti-CDC/N1MALVAC-1 antibodies can be used for
 CC detecting P. falciparum in biological samples.
 XX
 SQ Sequence 15 AA;
 QY Query Match 100.0%; Score 74; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 QYIKANSKFQIGTEL 15
 1 QYIKANSKFQIGTEL 15
 RESULT 14
 ID AAY44763 standard; peptide; 15 AA.
 XX AAY44763
 XX AAY44763;
 DT 04-MAY-2000 (first entry)
 DE Tetanus toxoid protein derived T-cell activating epitope P2.
 KM Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
 KM recombinant fusion protein cleavage; enzyme; protein polymer; foodstuff;
 KM antibacterial enzyme; surface glycoprotein; T-cell activating epitope;
 P2; tetanus toxoid; IPNV; Infectious pancreatic necrosis virus.
 XX Clostridium tetani.
 OS
 XX
 PN WO200004170-A1.
 XX
 PD 27-JAN-2000.

XX	14-JUL-1999;	99WO-CA00637.
PF		
XX	14-JUL-1998;	98CA-2237704.
FR		
XX	(UYBR-) UNIV BRITISH COLUMBIA.	
PA		
XX		
XX		
PI	Smit J;	
DR	WPI; 2000-182434/16.	
XX		
XX		
PT	Cleavage of Caulobacter produced recombinant fusion proteins useful for	
PT	producing vaccine peptides	
XX		
PS	Example 2; Page 16; 33pp; English.	
XX		
CC	The patent discloses a method for cleaving a recombinant fusion protein	
CC	which is produced by Caulobacter and consists of Caulobacter surface and	
CC	layer (S-layer) protein (containing the C-terminal secretion signal) and	
CC	a target protein heterologous to Caulobacter. The cleavage of target	
CC	protein from the S-layer protein is carried out under mild acid	
CC	conditions so that cleavage occurs at appropriate-proline dipeptide site	
CC	without solubilising the protein. The cleavage is accomplished while the	
CC	fusion protein is in an insoluble aggregate form which facilitates	
CC	purification of the protein. The method is useful for producing pure	
CC	proteins including recombinant human and animal therapeutic antibiotic	
CC	and vaccine peptides, enzymes, protein polymers, and antibacterial	
CC	enzymes for foodstuffs.	
CC	The present sequence is a T-cell activating epitope p2 derived	
CC	from tetanus toxoid protein. This sequence was fused to a DNA encoding	
CC	a fragment of infectious pancreatic necrosis virus surface glycoprotein	
CC	which is a vaccine candidate. This chimeric protein was in turn fused to	
CC	DNA encoding C. crescentus S-layer secretion signal (corresponds to the	
CC	C-terminal portion of the S-layer protein from amino acid 690 onwards and	
CC	contains native Asp-Pro site) for construction of a recombinant	
CC	fusion construct which is expressed in Caulobacter and then cleaved	
CC	to recover the vaccine candidate protein.	
XX		
SQ	Sequence	15 AA;
	Query Match	100.0%; Score 74; DB 21; Length 15;
	Best Local Similarity	100.0%; Pred. No. 3.9e-07;
	Matches 15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 QYKANSKFIGITEL 15	
DB	1 QYKANSKFIGITEL 15	
RESULT 15		
AAE11763		
ID	AAE11763 standard; peptide; 15 AA.	
AC	AAE11763;	
XX		
DT	18-DEC-2001 (first entry)	
XX		
DE	Clostridium tetani p2 epitope.	
XX		
KW	Amyloid protein; neuroprotective; nootropic; immunostimulant; vaccine;	
KW	Alzheimer's disease; anticonvulsant; gene therapy; Pick's disease;	
KW	antidiabetic; systemic amyloidosis; maturity onset diabetes; AIDS;	
KW	amyotrophic lateral sclerosis; Parkinson's disease; encephalopathy;	
KW	Huntington's disease; fronto-temporal dementia; p2 epitope.	
OS	Clostridium tetani.	
PN	MO200162284-A2.	
PD		
XX	30-AUG-2001.	
XX		
XX	19-FEB-2001; 2001WO-DK00113.	
XX		

PR 21-FEB-2000; 2000DK-0000265.
PR 01-MAR-2000; 2000US-186295P.
XX
PA (MEBI-) M & E BIOTECH AS.
XX Birk P, Jensen MR, Nielsen KG;
PI MPI; 2001-589796/66.
DR N-PSDB; AAD18755.
XX
PT In vivo down-regulation of amyloid protein for the treatment of
PT Alzheimer's, comprises presenting an amyloidogenic polypeptide or its
PT subsequence and/or at least one analogue of the amyloidogenic
PT polypeptide to the immune system -
XX
PS Example 3; Page 117; 120pp; English.
XX
CC The invention relates to a method for in vivo down-regulation of amyloid
CC protein such as beta amyloid (Abeta) in an animal, including human. The
CC method comprising presenting to the animal's immune system an
CC immunogenically effective amount of at least one amyloidogenic protein
CC or its subsequence and/or at least one analogue of the amyloidogenic
CC polypeptide. The amyloidogenic protein or its subsequence, and its
CC analogue is useful for the preparation of an immunogenic composition
CC comprising an adjuvant for down-regulating amyloid in an animal. They are
CC also useful in the treatment, prophylaxis or amelioration of Alzheimer's
CC disease or other diseases characterised by amyloid deposits. They are
CC also useful in the treatment of systemic amyloidosis, maturity onset
CC diabetes, Parkinson's disease, Huntington's disease, fronto-temporal
CC dementia, amyotrophic lateral sclerosis (ALS), Pick's disease and
CC prion-related transmissible spongiform encephalopathies. They are also
CC useful for inducing production of antibodies against an amyloidogenic
CC polypeptide. The present sequence is Clostridium tetani P2 epitope
CC related to the invention.
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 22; Length 15;
Best local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15
Db 1 QYIKANSKFIGITEL 15

Search completed: July 22, 2003, 08:08:15
Job time : 2.09218 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-149-163-12

Perfect score: 6810
Sequence: 1 MELALACRGGLLALLPPGA.....TFKGTPTAENPEYLGDVVP 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6734	98.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5941	87.2	1257	1 ERB2_RAT	P06494 ratus norv
3	5931.5	87.1	1254	1 ERB2_MESAU	O60553 mesocricetu
4	3158	46.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3136	46.0	1210	1 EGFR_MOUSE	O01279 mus musculu
6	2975.5	43.7	1308	1 ERB4_HUMAN	O15303 homo sapien
7	2957	43.4	1308	1 ERB4_RAT	O62956 ratus norv
8	2700.5	39.7	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2432.5	35.7	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2362.5	34.7	1339	1 ERB3_RAT	O62799 ratus norv
11	1967	28.9	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIER	P11273 avian eryth
15	1610	23.6	703	1 EGFR_CHICK	P13387 gallus galli
16	1297	19.0	1223	1 LTR2_CABEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculu
18	725	10.6	1363	1 ILPR_BRALA	O02466 branchiosto
19	708	10.4	1300	1 ILPR_MOUSE	O9W14 mus musculu
20	705.5	10.4	1372	1 INSR_MOUSE	P15208 mus musculu
21	705	10.4	1382	1 INSR_HUMAN	P06213 homo sapien
22	705	10.4	1383	1 INSR_RAT	P15127 ratus norv
23	702	10.3	1297	1 IRR_HUMAN	P14616 homo sapien
24	702	10.3	1607	1 MIRP_LYMST	O25410 lymnaea sta
25	699.5	10.3	1300	1 IRR_CAVPO	P14617 cavia porce
26	669	9.8	1477	1 HTK7_HYDAT	O25197 hydra atten
27	650	9.5	1367	1 IGR_HUMAN	P08069 homo sapien
28	640	9.4	1373	1 IGR_MOUSE	O60701 mus musculu
29	636.5	9.3	1370	1 IGR_RAT	O60701 mus musculu
30	634	9.3	1390	1 INSR_AEDAE	O93105 aedes aegypt
31	615	9.0	1246	1 INSR_DROME	P09208 drosophila
32	591.5	8.7	984	1 EPB1_CHICK	O07494 gallus galli
33	591	8.7	987	1 EPB4_HUMAN	P54760 homo sapien

34	589.5	8.7	977	1 EPB2_MOUSE	O03145 mus musculu
35	588	8.6	1114	1 RET_HUMAN	P07949 homo sapien
36	584.5	8.6	976	1 EPB2_HUMAN	P29317 homo sapien
37	577.5	8.5	987	1 EPB4_MOUSE	P54761 mus musculu
38	575.5	8.5	984	1 EPB1_RAT	P09759 ratus norv
39	575	8.4	985	1 EPB4_XENLA	O01571 xenopus lae
40	573.5	8.4	1053	1 FAK1_CHICK	O00944 gallus galli
41	571	8.4	902	1 EPBB_XENLA	O91736 xenopus lae
42	569.5	8.4	984	1 EPB1_HUMAN	P54762 homo sapien
43	569	8.4	1068	1 FAK1_XENLA	O91738 xenopus lae
44	567	8.3	757	1 HTK6_HYDAT	P53356 hydra atten
45	563	8.3	1052	1 FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6118653; PubMed=1003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
R Cussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Franke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
[3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2999967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor receptor gene and is amplified in a human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
[4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization.";
RL Genomics 15:426-429(1993).
-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11767; AAA35808.1; JOINED.
 DR EMBL; M11761; AAA35808.1; JOINED.
 DR EMBL; M11762; AAA35808.1; JOINED.
 DR EMBL; M11763; AAA35808.1; JOINED.
 DR EMBL; M11764; AAA35808.1; JOINED.
 DR EMBL; M11765; AAA35808.1; JOINED.
 DR EMBL; M11766; AAA35808.1; JOINED.
 DR EMBL; M11730; AAA5893.1; -.
 DR EMBL; M12036; AAA35978.1; -.
 DR EMBL; X03363; CAA27060.1; -.
 DR PIR; A25491; A25491.
 DR PIR; A24571; A24571.
 DR HSSP; P13362; IRGK.
 DR Genew; HGNC:3430; ERBB2.
 DR MIM; 164870; -.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Dkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Dkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_Dkinase; 1.
 DR SMART; SM00261; FU_3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 BY SIMILARITY.
 FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 I->V.
 FT VARIANT 655 655 /FTID=VAR_004077.
 FT VARIANT 655 655 I->V.
 FT CONFLICT 1170 1170 /FTID=VAR_004078.
 FT SEQUENCE 1255 AA; 137909 MW; 39E9DFDA04DCF962 CRC64;
 SQ
 Query Match 98.9%; Score 6734; DB 1; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
 1 MELALALRMGLLALLPRGAASVQCTGTDKRLPASPEHMLRLYLQGGCVVQGNL 60
 1 MELALALRMGLLALLPRGAASVQCTGTDKRLPASPEHMLRLYLQGGCVVQGNL 60
 61 ELTYLPTNASLFLQDIOEVGYVLIANQVQVPLQRLIVRGTOLEFEDYVALAVLNG 120
 61 ELTYLPTNASLFLQDIOEVGYVLIANQVQVPLQRLIVRGTOLEFEDYVALAVLNG 120
 121 DPLNNTTPVTGASFGGLREIQLRSLTEILKGVLIQNPQLCYDQTLMKQYIKANSKF 180
 121 DPLNNTTPVTGASFGGLREIQLRSLTEILKGVLIQNPQLCYDQTLMKQYIKANSKF 180
 121 DPLNNTTPVTGASFGGLREIQLRSLTEILKGVLIQNPQLCYDQTLMKQYIKANSKF 180
 121 DPLNNTTPVTGASFGGLREIQLRSLTEILKGVLIQNPQLCYDQTLMKQYIKANSKF 180
 181 IGIETLNRSPRACHCSPPCKSRGWSSEECOSLTRVAVGACRCARCPPTDCHCQC 240
 181 IGIETLNRSPRACHCSPPCKSRGWSSEECOSLTRVAVGACRCARCPPTDCHCQC 240
 181 IGIETLNRSPRACHCSPPCKSRGWSSEECOSLTRVAVGACRCARCPPTDCHCQC 240
 181 IGIETLNRSPRACHCSPPCKSRGWSSEECOSLTRVAVGACRCARCPPTDCHCQC 240
 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDFFESMPNPEGRTTGASCVTACP 300
 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDFFESMPNPEGRTTGASCVTACP 300
 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDFFESMPNPEGRTTGASCVTACP 300
 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDFFESMPNPEGRTTGASCVTACP 300
 301 YNYLSTVGSCTLVCPILHNOETVADGTORCEKSKPCARVCYGLGMEHREVRVATSAN 360
 301 YNYLSTVGSCTLVCPILHNOETVADGTORCEKSKPCARVCYGLGMEHREVRVATSAN 360
 301 YNYLSTVGSCTLVCPILHNOETVADGTORCEKSKPCARVCYGLGMEHREVRVATSAN 360
 301 YNYLSTVGSCTLVCPILHNOETVADGTORCEKSKPCARVCYGLGMEHREVRVATSAN 360
 361 IOEAGGCKTFGSLAFIPESFGDPASNTAPLOPEQLOVFETLEITGYLIAMPDSL 420
 361 IOEAGGCKTFGSLAFIPESFGDPASNTAPLOPEQLOVFETLEITGYLIAMPDSL 420
 361 IOEAGGCKTFGSLAFIPESFGDPASNTAPLOPEQLOVFETLEITGYLIAMPDSL 420
 361 IOEAGGCKTFGSLAFIPESFGDPASNTAPLOPEQLOVFETLEITGYLIAMPDSL 420
 421 DLSVFNQLQVIRGRILHNAGVSLTLQIGISIMLGRLSRLBAGSGALILHNTHLCFVNTV 480
 421 DLSVFNQLQVIRGRILHNAGVSLTLQIGISIMLGRLSRLBAGSGALILHNTHLCFVNTV 480
 421 DLSVFNQLQVIRGRILHNAGVSLTLQIGISIMLGRLSRLBAGSGALILHNTHLCFVNTV 480
 421 DLSVFNQLQVIRGRILHNAGVSLTLQIGISIMLGRLSRLBAGSGALILHNTHLCFVNTV 480
 481 PMDQLFNNPQOALLHTANRPEDECVEGGLACHOLCARGHCGPPTQCVCNCSQFLRGEC 540
 481 PMDQLFNNPQOALLHTANRPEDECVEGGLACHOLCARGHCGPPTQCVCNCSQFLRGEC 540
 481 PMDQLFNNPQOALLHTANRPEDECVEGGLACHOLCARGHCGPPTQCVCNCSQFLRGEC 540
 481 PMDQLFNNPQOALLHTANRPEDECVEGGLACHOLCARGHCGPPTQCVCNCSQFLRGEC 540
 541 VEEGRVYQGLPREVYNARHCLPCHPECOPONGSYTCGPEADQCACAHYDPPFCVARC 600
 541 VEEGRVYQGLPREVYNARHCLPCHPECOPONGSYTCGPEADQCACAHYDPPFCVARC 600
 541 VEEGRVYQGLPREVYNARHCLPCHPECOPONGSYTCGPEADQCACAHYDPPFCVARC 600
 541 VEEGRVYQGLPREVYNARHCLPCHPECOPONGSYTCGPEADQCACAHYDPPFCVARC 600
 601 PSQVTPDLSTYPIKFPDEEGACQPCPINCTHSCVLDLDDKCPAPORASPLTSIVSAVVG 660
 601 PSQVTPDLSTYPIKFPDEEGACQPCPINCTHSCVLDLDDKCPAPORASPLTSIVSAVVG 660
 601 PSQVTPDLSTYPIKFPDEEGACQPCPINCTHSCVLDLDDKCPAPORASPLTSIVSAVVG 660
 601 PSQVTPDLSTYPIKFPDEEGACQPCPINCTHSCVLDLDDKCPAPORASPLTSIVSAVVG 660
 661 ILLVTVLGVVFGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOAMRILKETEL 720
 661 ILLVTVLGVVFGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOAMRILKETEL 720
 661 ILLVTVLGVVFGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOAMRILKETEL 720
 661 ILLVTVLGVVFGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOAMRILKETEL 720

OY		721	RKVKYLGSACFTYVKGIMIPDGEVNVKI	PVALIKVLRENTSPKANKEIIIDELAYVMAGVSP	780
OY		721	RKVKYLGSACFTYVKGIMIPDGEVNVKI	PVALIKVLRENTSPKANKEIIDELAYVMAGVSP	780
OY		781	VYSRLIGICLTSTVOLVTQMLPBGCLLDHVENRGRLSODLNMCMOIAKMSYLEDDR		840
Db		781	VYSRLIGICLTSTVOLVTQMLPBGCLLDHVENRGRLSODLNMCMOIAKMSYLEDDR		840
OY		841	LVRDLAARNVLVKSPPNHVKITDFGLARLLDIDEFYHADGAKVIPKMALESILRRFPT		900
Db		841	LVRDLAARNVLVKSPPNHVKITDFGLARLLDIDEFYHADGAKVIPKMALESILRRFPT		900
OY		901	HOSDWASGVATWEMLTFCAGKPYDGI	PAIREIDPLEKEGRLPQPICITDVYMIWKCMM	960
OY		901	HOSDWASGVATWEMLTFCAGKPYDGI	PAIREIDPLEKEGRLPQPICITDVYMIWKCMM	960
Db		901	HOSDWASGVATWEMLTFCAGKPYDGI	PAIREIDPLEKEGRLPQPICITDVYMIWKCMM	960
OY		961	IDSECRPRPRELVSEFSRMARDPORFVVYIQNEDLGASPALDSTFYSRLLEDMDGLDYDA		1020
Db		961	IDSECRPRPRELVSEFSRMARDPORFVVYIQNEDLGASPALDSTFYSRLLEDMDGLDYDA		1020
OY		1021	EELYVPQCGFFCPDDAPGAGMWHHRHSSSTRSGGDLTLGLESEEBA	PRSPLAPSEG	1080
Db		1021	EELYVPQCGFFCPDDAPGAGMWHHRHSSSTRSGGDLTLGLESEEBA	PRSPLAPSEG	1080
OY		1081	AGSDVPFDGLMGCAKGLQSILPTHPSPLORRSEDPVPLPSETTGVAAPILTCSPQPEVY		1140
Db		1081	AGSDVPFDGLMGCAKGLQSILPTHPSPLORRSEDPVPLPSETTGVAAPILTCSPQPEVY		1140
OY		1141	NOPDVRPOPSPBREGPLPARBPAGATLERAKTISPCKNGVMDVAFGAVENPEYLTPO		1200
Db		1141	NOPDVRPOPSPBREGPLPARBPAGATLERAKTISPCKNGVMDVAFGAVENPEYLTPO		1200
OY		1201	GGAAPQHPPPPAFSAFNLYWDODDPBERGAPBSTFGTTPAENPEYLTGLDVPY		1255
Db		1201	GGAAPQHPPPPAFSAFNLYWDODDPBERGAPBSTFGTTPAENPEYLTGLDVPY		1255
Db		1201	GGAAPQHPPPPAFSAFNLYWDODDPBERGAPBSTFGTTPAENPEYLTGLDVPY		1255
<hr/>					
RESULT 2					
ERB2_RAT		ID_ERB2_RAT	STANDARD:	PRT: 1257 AA.	
AC		P06494;			
DT		01-JAN-1988 (Rel. 06, Created)			
DT		15-DEC-1998 (Rel. 37, Last sequence update)			
DT		15-JUN-2002 (Rel. 41, Last annotation update)			
DE		Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)			
DE		(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).			
GN		ERBB2 OR NEU			
OS		Rattus norvegicus (Rat).			
OC		Eukaryota; Metazoa; Chordata; Ratiata; Vertebrata; Euteleostomi;			
OC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX		NCBI_TaxID=10116;			
RN		[1]			
RP		SEQUENCE FROM N.A.			
RC		TISSUE=Neuroblastoma;			
RX		MEDLINE=86118662; PubMed=3945311;			
RA		Bargmann C.I., Hung M.-C., Weinberg R.A.;			
RT		"The neu oncogene encodes an epidermal growth factor receptor-related protein.";			
RL		Nature 319:226-230(1986).			
RN		[2]			
RP		SEQUENCE OF 852-905 FROM N.A.			
RC		TISSUE=Sciatic nerve;			
RX		MEDLINE=9122560; PubMed=2025425;			
RA		Lai C., Lemke G.;			
RT		"An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";			
RL		Neuron 6:691-704(1991).			
RN		[3]			
RP		STRUCTURE BY NMR OF 650-666.			
RX		MEDLINE=9215181; PubMed=1346763;			
RA		Gullik W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.			

RA	Newman R,Cumpton M.J., Stenberg M.J.E., Campbell I.D.;
RT	"Three dimensional structure of the transmembrane region of the proto-
RL	oncogenic and oncogene forms of the neu protein.";
CC	EMBO J. 11:43-48(1992).
CC	-I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC	ALTHOUGH NEUREGLINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC	ALPHA AND AMPHIREGULIN.
CC	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC	tyrosine phosphate.
CC	-I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC	THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I- PRM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC	RESIDUES (BY SIMILARITY).
CC	-I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X03362; CAA27059.1; ALT_INIT.
DR	PIR; A24562; TVRTNTU.
DR	HSSP; P11362; IFKC.
DR	InterPro; IPRO000494; EGFR_L_domain.
DR	InterPro; IPRO00719; Euk_Pkinase.
DR	InterPro; IPRO02174; Furin-like.
DR	InterPro; IPRO01245; Tyr_pkinase.
DR	InterPro; IPRO04019; YLP motif.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF01030; Recep_L_domain; 2.
DR	Pfam; PF02757; YLP_2.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	SMART; SMO0261; FU; 3.
DR	SMART; SMO0219; TYKc; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW	Proto-oncogene; Disease mutation.
FT	SIGNAL 1 21 POTENTIAL.
FT	CHAIN 1 21 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT	DOMAIN 22 1257 EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM 655 677 POTENTIAL.
FT	DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 159 369 CYS-RICH.
FT	DOMAIN 473 646 CYS-RICH.
FT	DOMAIN 722 989 PROTEIN KINASE.
FT	NP_BIND 728 736 ATP (BY SIMILARITY).
FT	BINDING 755 755 ATP (BY SIMILARITY).
FT	ACT_SITE 847 847 BY SIMILARITY.
FT	DISUPEID 196 205 BY SIMILARITY.
FT	DISUPEID 200 213 BY SIMILARITY.
FT	DISUPEID 221 228 BY SIMILARITY.
FT	DISUPEID 225 236 BY SIMILARITY.
FT	DISUPEID 237 245 BY SIMILARITY.
FT	DISUPEID 241 253 BY SIMILARITY.
FT	DISUPEID 256 265 BY SIMILARITY.
FT	DISUPEID 269 296 BY SIMILARITY.
FT	DISUPEID 300 312 BY SIMILARITY.
FT	DISUPEID 316 332 BY SIMILARITY.
FT	DISUPEID 335 339 BY SIMILARITY.
FT	DISUPEID 513 522 BY SIMILARITY.
FT	DISUPEID 517 530 BY SIMILARITY.
FT	DISUPEID 533 542 BY SIMILARITY.
FT	DISUPEID 546 562 BY SIMILARITY.
FT	DISUPEID 565 578 BY SIMILARITY.

```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 87.2%; Score 5941; DB 1; Length 1257;
Best Local Similarity 87.2%; Pred. No. 6,1e-309;
Matches 1096; Conservative 51; Mismatches 108; Indels 2; Gaps 2;

QY 1 MELALCRMGILLALLPFGAASITQVCTGDMKRLPASPEHLDMLBHLVQGCQVQGNL 60
D 1 MELALCRMGFLALLPFGIAGTQVCTGDMKRLPASPEHLDMLBHLVQGCQVQGNL 60
QY 61 ELTYLPNASLSFLQDIQEVQGYVLIAHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
D 61 ELTYVPANASISFLQDIQEVQGYVLIHNOVQVPLQRLIRVGTQLFEDNYALAVLDNR 120
QY 121 DPLNNTTPTV-GASPGGLRELQRLSTEILKGGVLIQBNPOLCYODITLMDQYIKANSK 179
D 121 DPLNNTTPTVGRIGRELQRLSTEILKGGVLIQBNPOLCYODITLMDQYIKANSK 180
QY 180 FIGITELNRSRACHPCSPMGKSRGWESSESDCSLTTRVACAGCARCKGFLPTDCHEQ 239
D 180 APVDIDNRSRACHPCSPMGKSRGWESSESDCSLTTRVACAGCARCKGFLPTDCHEQ 240
QY 240 CAAGCTGPKASDCLACHFNHSGICEIHCPLVNTNTDTESMPNPEGRTYFGASCYTAC 299
D 241 CAAGCTGPKASDCLACHFNHSGICEIHCPLVNTNTDTESMPNPEGRTYFGASCYTAC 300
QY 300 PYNVLSTDVSGCTLVCPHNOEYTAEDGTORCEKSPKARVCYGLMEHLREVRATSA 359
D 301 PYNVLSTEVSGCTLVCPHNOEYTAEDGTORCEKSPKARVCYGLMEHLREVRATSD 360
QY 360 NIOEFACCKKIFGSLAFPEPSFDDPASNTAPLOPELOVEFETLEITGYLYISAMPDSL 419
D 361 NVGEFDCCKKIFGSLAFPEPSFDDPASNTAPLOPELOVEFETLEITGYLYISAMPDSL 420
QY 420 PDLVSFONLOVIRGRIIHNAYSLTLOGIGISWGLRLSRLSGGLALIHNTHLCFVHT 479
D 421 PDLVSFONLRIIRGRIIHNAYSLTLOGIGISWGLRLSRLSGGLALIHNTHLCFVHT 480
QY 480 VPMDOLEFRNPHOALLHTANRPEDE-CVGEGLACHQLCARHGCMWGPRTQCCNCSQPLRGQ 538
D 481 VPMDOLEFRNPHOALLHTANRPEDE-CVGEGLACHQLCARHGCMWGPRTQCCNCSQPLRGQ 540
QY 539 ECVEECRVLQGLPREYVNAHCLPCHPECOFONGSVTCFGEADQVACAHYKDPFCVA 598
D 541 ECVEECRVMKGLPREYVSDKRLCPCHPECOFONGSVTCFGEADQVACAHYKDPFCVA 600
QY 599 RCSSGVKPDLSYPIKMFPEBEGACQCPINCHSCVDLDDKGPAPQASPLTSYSAV 658
D 601 RCSSGVKPDLSYPIKMFPEBEGACQCPINCHSCVDLDDKGPAPQASPLTSYSAV 660
QY 659 VGLILVVLGVFGIILKRRQOKIRKTYMRLLQETLVEPLTPSGAMPQAOQRILKET 718
D 661 VGLILVVLGVFGIILKRRQOKIRKTYMRLLQETLVEPLTPSGAMPQAOQRILKET 720
QY 719 ELRKRVLSGAGFTYVKGIMIPDGENVKI PVAIKVLRENTSPKANKEILDEAYVMAVG 778
D 721 ELRKRVLSGAGFTYVKGIMIPDGENVKI PVAIKVLRENTSPKANKEILDEAYVMAVG 780
QY 779 SPVSVRLGICLTSTQVLTQVQMPYGCGLDHVNRGRLOSLDLNMCQIAGKMSYLED 838

```

```

D 781 SPVSVRLGICLTSTQVLTQVQMPYGCGLDHVNRGRLOSLDLNMCQIAGKMSYLED 840
QY 839 VRLVHRLAARNLVKSPNHYKTDPEGLARLDIDETEVHADGKVPKIMALESLIRRR 898
D 841 VRLVHRLAARNLVKSPNHYKTDPEGLARLDIDETEVHADGKVPKIMALESLIRRR 900
QY 899 FTHQSDWVSQVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVWC 958
D 901 FTHQSDWVSQVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVWC 960
QY 959 WMIDSECRPFRELVSFSSMARDPORFVYIQNEDICPASPLDSTFYRLSEDDMDGLV 1018
D 961 WMIDSECRPFRELVSFSSMARDPORFVYIQNEDICPASPLDSTFYRLSEDDMDGLV 1020
QY 1019 DAEEYLVPOGFCPPDPAPAGCMVHRHSSSTRSGGDTLTGLESEEPAPSPAPLS 1078
D 1021 DAEEYLVPOGFCPPDPAPAGCMVHRHSSSTRSGGDTLTGLESEEPAPSPAPLS 1080
QY 1079 EGAGSDVFDGDLGMAKGLQSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1138
D 1081 EGAGSDVFDGDLGMAKGLQSLPHDLSPLQRYSEDPVPLPSETDGYVAPLTCSPQE 1140
QY 1139 YVNOQDVRPQPSPEBGLPAPAPAGATLERAKTSLSPGKGVKVDVAFGAVENPEYLT 1198
D 1141 YVNOQEVQPPPLPPEBGLPAPAPAGATLERAKTSLSPGKGVKVDVAFGAVENPEYLT 1200
QY 1199 POGGAAPQPPPAFSPAFNLYWDDPREBQAPSTFTGTPAENPEYLGIDLVV 1255
D 1201 PREGTASPPHSPAFSPAFNLYWDDPREBQAPSTFTGTPAENPEYLGIDLVV 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D16295; BAA03801.1; .
CC HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1254
FT DOMAIN 22 1254
FT TRANSMEM 653 675
FT DOMAIN 676 1254
FT DOMAIN 158 368
FT DOMAIN 472 644
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT DISULFID 195 204
FT DISULFID 199 212
FT DISULFID 236 244
FT DISULFID 255 262
FT DISULFID 255 264
FT DISULFID 268 295
FT DISULFID 299 311
FT DISULFID 315 331
FT DISULFID 334 338
FT DISULFID 511 520
FT DISULFID 515 528
FT DISULFID 531 540
FT DISULFID 544 560
FT DISULFID 563 576
FT DISULFID 567 584
FT DISULFID 587 596
FT DISULFID 600 623
FT DISULFID 626 634
FT DISULFID 642 642
FT MOD_RES 1139 1139
FT MOD_RES 1247 1247
FT CARBOHYD 68 68
FT CARBOHYD 125 125
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 658 658
FT VARIANT 659 659
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;
Query Match 87.1%; Score 5931.5; DB 1; Length 1254;
Best Local Similarity 86.9%; Pred. No. 2e-308;
Matches 1091; Conservative 59; Mismatches 104; Indels 1; Gaps 1;

QY 61 ELTYLPNMSLSLFLQDIQEVQGVLLAHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPANATLSFLQDIQEVQYMLAHSGVRRVPLQRLIRVGTQLFEDNYALAVLDNR 120
QY 121 DPLANTTPVYGASPGGIREQLSLSTELLKGVLIQNRPNLCTODTTLKMDQYIKANSKF 180
DB 121 DPLDNTTATGATGPRGELRELQLSLSTELLKGVLIQNRPNLCTODTTLKMDQVFRKXNQLA 180
QY 181 IGITELNRSRACHPCSPMCKGRSGESSSEDCSLFTVAVAGCARCKGPPPTCCHEQC 240
DB 181 PVIDIDNRSRACHPCAPACKDNHCCKGASPEDCQTLLCTIAPRAVPAARALPDDCCHEQC 240
QY 241 AAGCTGPKSIDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFCASCVTACP 300
DB 241 AAGCTGPKSIDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFCASCVTACP 300
QY 301 YNLTSTEVGSCITVCPILNNOEVTAEEDGTQCEKCSKPCARVCYGLGMEHLRGARALITSAN 360
DB 301 YNLTSTEVGSCITVCPILNNOEVTAEEDGTQCEKCSKPCARVCYGLGMEHLRGARALITSAN 360
QY 361 IOFACCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVETLEITGYLYISAMPDSL 420
DB 361 IOFACCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVETLEITGYLYISAMPDSL 420
QY 421 DLSVFNQLQVIRGRIHANGAYSLTLQGLGISWLGSLRELGSGLALIHNTHLCFVHTV 480
DB 421 DLSVFNQLQVIRGRIHANGAYSLTLQGLGISWLGSLRELGSGLALIHNTHLCFVHTV 480
QY 481 PMQULFRNPHQALLHFNANRPEDECVEGLACHQLCARHGWGPPQCVNCSQFLRGQEC 540
DB 481 PMQULFRNPHQALLHFNANRPEDECVEGLACHQLCARHGWGPPQCVNCSQFLRGQEC 540
QY 541 VEECRVLOGIPREYVNAHCLPCHPECPONSGSVTFCGPEADOCVCAHYKXDPFCVARC 600
DB 541 VEECRVLOGIPREYVNAHCLPCHPECPONSGSVTFCGPEADOCVCAHYKXDPFCVARC 600
QY 601 PSQVKEPDLSTYMPWKPEPDEGACQPCPINCTHSCVDLDKGCPEAERASPLTISVAVNG 660
DB 601 PSQVKEPDLSTYMPWKPEPDEGACQPCPINCTHSCVDLDKGCPEAERASPLTISVAVNG 660
QY 661 ILLVVLGVVFGILLRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNOQMRLKTEL 720
DB 661 ILLVVLGVVFGILLRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNOQMRLKTEL 720
QY 721 RKVKVLGSAFGVYVYGIWIPGSENVKIPVAILVLEENSPKXKELIDEAAYMAGGSP 780
DB 721 RKVKVLGSAFGVYVYGIWIPGSENVKIPVAILVLEENSPKXKELIDEAAYMAGGSP 780
QY 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENHGRGLSODLLNMCQIAKSGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENHGRGLSODLLNMCQIAKSGSYLEDVR 840
QY 841 LVHRDLAANVLYKSNHVKITDFGLARLLIDETEHADGKVPIKMALESILRRFT 900
DB 841 LVHRDLAANVLYKSNHVKITDFGLARLLIDETEHADGKVPIKMALESILRRFT 900
QY 901 HOSDWVSVTVTWEIMTFCAPYDGI PAEIPDLLEKGRLLPQPCITTDVYIMKXCM 960
DB 901 HOSDWVSVTVTWEIMTFCAPYDGI PAEIPDLLEKGRLLPQPCITTDVYIMKXCM 960
QY 961 IOSECRPRELIVSESRMARDPORFVYIQNEDLGASPLDSTFTFYSLLLEDDMGVLVDA 1020
DB 961 IOSECRPRELIVSESRMARDPORFVYIQNEDLGASPLDSTFTFYSLLLEDDMGVLVDA 1020
QY 1021 EBYLYVPOGFCPPDPAAGAGVHHHRSSSTRSGGDLTTLGSPSEEAAPRSLAPASEG 1080
DB 1021 EBYLYVPOGFCPPDPAAGAGVHHHRSSSTRSGGDLTTLGSPSEEAAPRSLAPASEG 1080
QY 1081 AGSDVFDGGLGMAKGLQSLPETHDPSPLQRISEDPVTPLPSETGTGYVAPLTCSPQPEV 1140
DB 1081 AGSDVFDGGLGMAKGLQSLPETHDPSPLQRISEDPVTPLPSETGTGYVAPLTCSPQPEV 1140

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP ASN-528.
RA MEDLINE=96398132; PubMed=8962717;
RX Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT "Analysis of the glycosylation patterns of the extracellular domain of
RT the epidermal growth factor receptor expressed in Chinese hamster
RT ovary fibroblasts";
RL Growth Factors 13:121-132(1996).
RN [17]
RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP ASN-603.
RX MEDLINE=20198209; PubMed=10731668;
RA Sato C., Kim J.-H., Abe Y., Saico K., Yokoyama S., Kohda D.;
RT "Characterization of the N-oligosaccharides attached to the atypical
RT Asn-X-Cys sequence of recombinant human epidermal growth factor
RT receptor";
RL J. Biochem. 127:65-72(2000).
RN [18]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=98225196; PubMed=9556602;
RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT "Disulfide bond structure of human epidermal growth factor receptor";
RL J. Biol. Chem. 273:11150-11157(1998).
RN [19]
RP REVIEW.
RX MEDLINE=87297456; PubMed=3039909;
RA Carpenter G.;
RT "Receptors for epidermal growth factor and other polypeptide
RT mitogens";
RL Annu. Rev. Biochem. 56:881-914(1987).
CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
CC involved in the control of cell growth and differentiation.
CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC secreted.
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/
CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC expressed in ovarian cancers.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00588; CA25240.1; -
DR EMBL: U95089; AAB53063.1; -
DR EMBL: U48722; AAC50802.1; -
DR EMBL: U48723; AAC50804.1; -
DR EMBL: U48724; AAC50796.1; -
DR EMBL: U48725; AAC50797.1; -
DR EMBL: U48726; AAC50798.1; -
Query Match 46.4%; Score 3158; DB 1; Length 1210;
Best Local Similarity 49.8%; Pred. No. 7.8e-161;
Matches 631; Conservative 176; Mismatches 349; Indels 112; Gaps 23;
QY 11 LLLALLPPGAA--STGVCTGTDMLRLPASPETHLDMLRYOGCCOVVQGNLELYLPTN 68
DB 14 LLAALCPASRALEKKEKVCQSTNSKLTQVLGFEDHPLSLQMFNNCEVVLGNLEITYVORN 73

QY 69 ASLSFLDIDIEVOGVYLIANOVROVPLQRLIRVGTOLFEDNYALAVLNDGPDLANNTP 128
DB 74 YDLSEFLTKIDIEVAGVYLIANTVERIPLENQIIRGNMYENSVALAVLSNYD----- 126
QY 129 VTGASPEGLELQRLSRITELTKGVLIQRNPOLCYQDTILMKQOYIANSFICITEL-- 186
DB 127 ---ANKTKLELPRNIOELIHGAVRPSNNPALCNVSIQWRD---IVSSPFLSNMGMD 180
QY 187 -NRSRACHPCSPCKSGRSGESEDQSLTRVCAGGCA-RCKGPLPTDCCHQCAAGC 244
DB 181 QNHLGSCQKDPSCGPNCSWGAGEENCKTKITICAQCCGRCKRSPSDCHQCAAGC 240
QY 245 TGRPHSDCLALFHNNHGICELCPALVTYNTDFESMPNPEGRTYTGASCVTAQPNYL 304
DB 241 TGPRESDCIVCRKFRDPCATCKDCPMLNPTTYQMDVNVDEKYSFGATCVKCKPRNVV 300
QY 305 STDVSGTLVCPPLNDEVTADGTQREKSKPRAPVCYGLGMEHLEVRVTSANIQEF 364
DB 301 VTDHGSVCVRACGADSYEM-EDGVKCKCKEGPCRKVCNGIGIEFDSLSINATNIKHF 359
QY 365 AGCKITFGSLAPLPESFPDGPASNTAPLOPEQLOVFTEITGYLYISAMPDLPDLSV 424
DB 360 KNTCTISGDLHIIPVAFRGSFTHTPPLDQELDILKTVKELITGFLIQAHPENRTDLHA 419
QY 425 FQNLQVIRGRILHNQAVSLTLOGLISWLGRLSRLSGSLALIHNTHLCFVHTVPMDQ 484
DB 420 FENLEIRGRTKQHGQSLAVSLNITSLRLSKETISDDDVIIISGNKLCYANTIMWK 479
QY 485 LFRNPHQALHTANRPDEBCEVGBGLACHQLCARHCCKGPPPTQCNCVSQFLRQECVBC 544
DB 480 LFGSGQKTKIISNGENSCAKAGVCHALCSPEGCGPBPDPVCSGRNVRGREGVDKC 539
QY 545 RVLOGLPREVYNNRHLCPHPECOPOPNQSVTCGPRADQCVACAHYDPPEPCVRCPSGV 604
DB 540 NLGEPRERVENSECIQCHPECLPQAMNLTCTGRPDNCTIQAHYIDGPRCVATPCAGV 599
QY 605 KPDLSTYPIWKPFDEBACOPCPINCTHSCVDLDDKCPAEBASPLTISAVAG---I 661
DB 600 MGENNTL-VKRYADAGVCHLCHPNCTYCGTGGLBCPLNGPRIP--STATGVNGLLL 656
QY 662 LTVVLGVVFGIILIKRQOKIRKRYTMRLLIQETELVEPLTPSGAMPQAOQRILKETELR 721
DB 657 LTVVALGIG--LFMRRIHVRRKRLRLIQEBELVEPLTPSGEAPQALRLIKETEFK 713
QY 722 KVVVLGSGAGTVYKGIWIPDGENVKIPVIXYLRENTSKANKEILDEAYVYVAGVSPY 781
DB 714 KIVVLGSGAGTVYKGIWIPGEKVKIPVIXKLEKATSPKANKEILDEAYVYVAVSDNP 773
QY 782 VSRLLGICLTSTVQLVTLMPYGCILDHVENRGRGLSQDPLNMCMQIAGMSYLEDVRL 841
DB 774 VCRLLGICLTSTVQLTQLMPFGCLLDYVBEHNDIGSYLLMKVCQIAGMNTLEDRL 833
QY 842 VHRDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADGKYPIKMALESILRRRFT 901
DB 834 VHRDLAARNLVKTPQGVKITDFGLAKLGAEEKYVIEAGKVPDKMALESILHRYTH 893
QY 902 QSDVMSYGVYVWMLMTPGAPYDGIIPAREIPDLLEKGERLPORPICITDYVMIMVKMMI 961
DB 894 QSDVMSYGVYVWMLMTPGSKRPYDGIIPASEISSLLEKGERLPORPICITDYVMIMVKMMI 953
QY 962 DSECRPRELVESEFARMARDPORFVYIQ-NEDLGAPSPDSTFFYSRLBEDDDMGDLVDA 1020
DB 954 DADSRPKRRLIIEFSKWARDPORVIVIQODERMHLPSFPDSNFYRLALMBEDMDVDVDA 1013
QY 1021 EBYLVPOQGFPCPDPAAGAGVYHRRSSSTRSGGDLITLGLPSEEAAPSLPABSG 1080
DB 1014 DEYLIVPOQGF-----SSPSTRPPLSLSS 1039
QY 1081 AGSDVPDGLGMAAGLQSLPTHDPSPLORESDPVLPSET--DGYVAPLTCGQPE 1138
DB 1040 ATSN--NSTVACIDRNGLOSCPIKEDSPLORYSSDPTGALTDSIDDTFL-----FVPE 1091

QY 1139 YVNOPDVPRPSPREGPLPAAPAGATLERAKTLSPGKNGVWVDFAGGAVENPEYL- 1197
 DB 1092 YVNO-SVPRKRPASVQNVPHNQPLNP-----ASRDPHYQD--PHTAVGNPEYLN 1140
 QY 1198 TPOGGAAPQHPAPPASPAADNLTYWDQ-----DP-----PERGAPPSTFGTPT 1241
 DB 1141 TVO-----FTCVNSTPDSPAHMAQKSHQISLDNPPYQODFPFKKAKPFGIS- 1190
 QY 1242 TAENPEYL 1249
 DB 1191 TAENAEYL 1198

RESULT 5

EGFR_MOUSE
 ID EGFR_MOUSE STANDARD; PRT: 1210 AA.

AC Q01279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yayon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (bek/KGFR) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbe M.L.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170966; PubMed=8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase.";
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9122866; PubMed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 binding site.";
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Eisinger D.P., Serrero G.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 dimerization, internalization of the EGF-receptor complex,
 induction of the tyrosine kinase activity, stimulation of cell DNA
 synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>
 or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; X78987; CAAS5587.1; -;
 DR EMBL; U03425; AAA17899.1; -;
 DR EMBL; X59698; CAA42219.1; -;
 DR EMBL; L06864; AAA53029.1; -;
 DR EMBL; Z12608; CAA78249.1; -;
 DR HSSP; P11362; IFGK.
 DR MGD; MGI:95294; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Prodom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; Fu; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 647
 FT TRANSMEM 648 670
 FT DOMAIN 671 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1028 1071
 FT DOMAIN 714 981
 FT NP_BIND 720 728
 FT BINDING 747 747
 FT ACT_SITE 839 839
 FT DISULFID 190 199
 FT DISULFID 194 207
 FT DISULFID 215 223
 FT DISULFID 219 231
 FT DISULFID 232 240
 FT DISULFID 236 248
 FT DISULFID 251 260
 FT DISULFID 264 291
 FT DISULFID 295 307
 FT DISULFID 311 326
 FT DISULFID 329 333
 FT DISULFID 332 345
 FT DISULFID 506 515
 FT DISULFID 510 523
 FT DISULFID 526 535
 FT DISULFID 539 555
 FT DISULFID 558 571
 FT DISULFID 582 591
 FT DISULFID 595 617
 FT DISULFID 620 628
 FT DISULFID 624 636
 FT MOD_RES 680 680
 FT MOD_RES 1092 1092
 FT MOD_RES 1092 1092
 CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```

FT  MOD_RES  1110  1110  PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT  MOD_RES  1172  1172  PHOSPHORYLATION (AUTO-) (BY SIMILARITY)
FT  MOD_RES  1197  1197  PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT  MOD_RES  1197  1197  (BY SIMILARITY)
FT  CARBOHYD  128  128  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  175  175  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  196  196  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  352  352  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  413  413  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  444  444  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  528  528  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  568  568  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  603  603  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  623  623  N-LINKED (GLCNAC) (POTENTIAL)
FT  CARBOHYD  19  19  C -> S (IN REF. 2)
FT  CONFLICT  539  539  C -> W (IN REF. 5)
FT  CONFLICT  991  991  L -> F (IN REF. 4)
FT  CONFLICT  1116  1117  HP -> DR (IN REF. 6)
SQ  SEQUENCE  1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

```

```

Query Match 46.0%; Score 3136; DB 1; Length 1210;
Best Local Similarity 49.6%; Pred. No. 1,1e-159;
Matches 629; Conservative 171; Mismatches 366; Indels 102; Gaps 21;

```

```

QY  11 LLLALLPAGA--STQCTGTDMKLRPASBETLDMRLHYQCCQYVQGNLELTYPTN 68
DB  14 LLLALLPAGA--STQCTGTDMKLRPASBETLDMRLHYQCCQYVQGNLELTYPTN 73
QY  69 ASTSLFDIDIEVQGYVLIANNVOYRPLQRLRIYRGTQLFEDNVALVLDNGDLNNTTP 128
DB  74 YDSFLKTIQVAGYVLIANNVOYRPLQRLRIYRGTQLFEDNVALVLDNGDLNNTTP 124
QY  129 VTGASRGLRELQRLSLTEILKGGVLIQRNPOLCYODTIIMKQYIKANSKFIGITELNR 188
DB  125 -YGTNRIGRELQRLSLTEILKGGVLIQRNPOLCYODTIIMKQYIKANSKFIGITELNR 183
QY  189 SRACHRECPMKCKSRCKGSESSDQSLTRTVYAGGCA-RCKGLPTDCCHEQCAAGTGP 247
DB  184 PSSCPRKCDPSCPGSCGGEENCQKLTIKIICAOQCHRCGRGRPSDCCHQCAAGTGP 243
QY  248 KHSDDLACLFHNSGICELCHLPALVTNTDTFESMPREGRYTPGASCAVACPNTYSTD 307
DB  244 RESDCLVCKQFQDEATCKOTCPRLMLNPTTYQMDVNPBGKYSIGATCVKCCPRNYYVTD 303
QY  308 VGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVYAGMEHLREVAVTSANTIOEPAGC 367
DB  304 HGSVCVACGPDYEV--BEDGIRKCKKCDGCRKVCNGSIGIGEPFDLISINATNIKHKYC 362
QY  368 KRTFGLAPLPESFDGPANTAPLOEQLQVETLEITGYLYISAMPDLSPLDLSVFON 427
DB  363 TALSGLHLPLVAFKGSFRTPLDRELEILKTVEITGFLIQAMPDNTDLHAFFEN 422
QY  428 LOYIRGRILHNGVYSLTLOGLSISWLSRLSRLSGLALIHNTTHLCFHTYPMWDLFR 487
DB  423 LEIIRKRTKQHGFSLAVNGVLTSLGRLSKLSIDSDVLTISGRNLCYANTNTIMWKLFG 482
QY  488 NPHQALLHTANRDEDECVSEGLACHOLCARGHMGCPPTQVNCQOFLRGQECVEECRLV 547
DB  483 TPQKTKIKMNRKAEKCKAVNHCNPLCSECGMGPRPCVGCQVNSRRECEVEKNIL 542
QY  548 QGIPREYVNAARHLPCGHECQOPONGSVTCGPEADQCVACAHYKDPFCVACRPSGVKP 607
DB  543 EGPREFEVNSECIQCHPECLPQAMNITCTGRPDNCTIQCAHYIDGPHCVKTCPAGIMGE 602
QY  608 LSTYPMKFPDEGACOPCPINCTHSCVDLDDKCAPAQASPLTISVAVNGILLVVL 667
DB  603 NNTL-VMKADANNVCHLANCTYGCAGGLOGCEWPSGPKIPSTANTIGVGLLFTIV 661
QY  668 GAVFEG-LIKRQOKIKRYTMRRLQETLEVBPLTSGAMPNOQAKILKELTRKVKYL 726
DB  662 -VALGIGLPMRRHRIYKRTLRRLQEBRELVEPLTSGEAPNOAHLRIKETEFLKIKYL 720
QY  727 GSGAGFTVYKGIWIPGENVKIPVALKVLRENTSPKANKEILDEAYVMAGVSPYVSRL 786

```

```

DB  721 GSGAGFTVYKGIWIPGENVKIPVALKVLRENTSPKANKEILDEAYVMAGVSPYVSRL 780
QY  787 GICLSTVQVLTQMPGICLLDHYENRGRIGSODLNMCMQIKKMSYLEDLVLYHDL 846
DB  781 GICLSTVQVLTQMPGICLLDHYENRGRIGSODLNMCMQIKKMSYLEDLVLYHDL 840
QY  847 AARNVLYSPNHVITDPGLARLIDIDETEYHAQGVKPIKMALESLIRRRFTQSDVM 906
DB  841 AARNVLYKTPQHVKITDPGLAKLIGAEKEKHNACGKVPKIMMALESLIRHTQSDVM 900
QY  907 SYGTWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYIMVCMIDSECR 966
DB  901 SYGTWELMTFGSKPYDGIIPASDISILKEXGERLPQPICTIDVYIMVCMIDADSR 960
QY  967 PREFLWSEFSRMARDQRFVLTQ-NEDLQASPLDSTFRYSLLEDDMGVLVDAEVLV 1025
DB  961 PREFLWSEFSKMRDQRFVLTQDERMHLPSPTDINFYALNDEEDMEVVADEFLI 1020
QY  1026 POGGFCPPDPAPGAGVHHRHRSSTRSGGDLTLGLEPSEEARPSPLAPSEGAGSDV 1085
DB  1021 POGGFCPPDPAPGAGVHHRHRSSTRSGGDLTLGLEPSEEARPSPLAPSEGAGSDV 1045
QY  1086 FDGDLGMAAGLQSLPTHPSPLOYSDEPTVLPSET--DGVAPLTCSPOBEVYNQ 1143
DB  1046 ---NSTVACINRNSCRVKEDAFLORYSSDPTGAVTEDNIDDAFL-----PVPEVYNQ- 1095
QY  1144 DVRRQPSRREGPLPARPACATLERAKTISPGKGVKQVDFAFAGAVENPEYL-TPOGG 1202
DB  1096 SVRRPAGSVQNPYHNPQPLHP-----APGDLHYQN--PHSNVAGPVEYNTAQO-- 1143
QY  1203 AAPQHPPPAFSPAFDLVYWDQ-----DP-----PERGAPSTFKGPTENP 1246
DB  1144 -----PTLSSGFSNPALMIQKSHQMSLNDPVDYQDFFPKTKVNGIFKG-PTAENA 1195
QY  1247 EYLGDLVP 1254
DB  1196 EYLRVAPP 1203

RESULT 6
ERB4 HUMAN STANDARD; PRT; 1308 AA.
ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8983326;
RA Plowman G.D., Culicou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.,
RT "ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Plowman G.D.,
RA Klagsbrun M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester."
RL J. Biol. Chem. 272:26761-26768(1997).

```



```

Db      536 CNLYDGEFREFENSGICVECDPOCEKMEBDGLTGHGROPNDCTKSHFKDGNVCEKCPD 595
Qy      603 GVKPDLSTYMPJWKFPEDEGACQPPINCTHSCVLDKGC-----PAEQRASPL 651
Db      596 GLQGANF--IFKYADPDRECHPCHPNCOTQCNPTSHDICYPTWGTGSHLTPQAR--TPL 652
Qy      652 TSIVSAV--GILLVVLVGVGILIKRRQOKIRKTYTRRLLQFELVEPLTPSGAMPQA 710
Db      653 --IAAGVIGGLFIVLIVGLTFVAVYRKRKSIK-KRALRRFL-ETELVEPLTPSGATPQA 708
Qy      711 QMRILKETELRKVAVLGSAGFVYKGIWIPDGENKIPVAIKYLRNTSPKAKEILDE 770
Db      709 QLRILKETELRKVAVLGSAGFVYKGIWIPDGENKIPVAIKYLRNTSPKAKEILDE 768
Qy      771 AYVAVGSPVYSRLGICLTSTVQVLTQMPYSCLLDHYENRGLSGODLIMWCQIA 830
Db      769 ALIMASMDHPLVRLGLVCLSPITQVLTQMPHSCCLLEVYEHKDNIGSQLLMWCQIA 828
Qy      831 KGMVYLEDVRLVHRDLAARVAVLVSPPHVKITDGLARLLDIDETEVHADGKPIKMA 890
Db      829 KGMVYLEDVRLVHRDLAARVAVLVSPPHVKITDGLARLLDIDETEVHADGKPIKMA 888
Qy      891 LESILRRFTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPPICTID 950
Db      889 LECIHVRKFTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPPICTID 948
Qy      951 VVWIMVYKCMVIDECPREFELVSEFSRMAADPQRFVIONED-LGPASPLDSTFYKSL 1009
Db      949 VVWIMVYKCMVIDECPREFELVSEFSRMAADPQRFVIONED-LGPASPLDSTFYKSL 1008
Qy      1010 EDDMDGLVDAEELVYPOQGFCDPAPAGAGVHHHRRSSSTRSGGDLTLGLEBEE 1069
Db      1009 DEEDLEDDMDAEELVY-QAFLNIPPP-----YTSHARIDSNSS-----EIGHSPRAY 1056
Qy      1070 APRS-----PLAP-SEGAGSDVPDQGLDMGAKGLOS 1100
Db      1057 TPMGNGFVVRDGGFAAGGVSVYRAPSTIPEAPVAGQATAIIFDSCNGTLKRPVA 1116
Qy      1101 LPHDSEPLQRYSEDPVPLPS-----ETDGVAVLTCSPQPEYVNOQDVRPQSPSR 1153
Db      1117 PHVQDSSTQRYSDADPTVFAPERSPRGELDEEGYMTMRDKPQGEVYNPVE----- 1167
Qy      1154 EGPLPARPAGATLERAKTISPKNGVYKQVFAFGCAVENPEYITPPGGAAPQDHPRA- 1212
Db      1168 ENDFVSR-----KNGDLO-----ALDPEYHNSNG-----PKAE 1199
Qy      1213 -----FSPAFDNLVYWDQDPPERGA--PPSTF 1237
Db      1200 DEVVNEPLVNTANTLGAKEYLKNILSMPEKAKKAFNDPDMYWNHSLPPRSTLQHPDYL 1259
Qy      1238 KGTPT-----AENPEYL 1249
Db      1260 QEYSTKYFYKQNGRIRPIVAENPEYL 1285

```

```

RESULT 7
ERB4 RAT
ID ERB4 RAT STANDARD; PRT; 1308 AA.
AC 062956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269 (1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704 (1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97164212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF041838; AAD08899.1; -.
DR EMBL; U52531; AAC51051.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyRKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB096A08B41 CRC64;
```

Query Match 43.4%; Score 2957; DB 1; Length 1308;
Best Local Similarity 45.1%; Pred. No. 4.2e-150;
Matches 608; Conservative 191; Mismatches 388; Indels 160; Gaps 28;

```
QY 1 MEIA-ALCRWGLL--ALLPFGAASVCTGTDMKRLPASPTHLDMLRHLVGGCQVQ 57
D 1 MKLATGLMWGSLVLAARTVQPSASQSVCACTENKLSLSDLEQVYALRKYENCENV 60
QY 58 GNLELYLPTNASSLFIODIOEVQYVLIANQVROPLORLIVRGTOLEFEDNYALVL 117
D 61 GNEIETSIENHRDLSFLRSIREVGYVALNQFRYPLLENLRIRIRGKYLEDEYALAI 120
QY 118 DNGDPLNNTPTVYGASPGGLREQLASLTTELKGVLIQRNPOLCYQDDTIMKOYIKAN 177
D 121 LNRKQDNF-----GLOELGKLTTELINGVYVDQNKFLCYADTIHMUDI VNPW 171
QY 178 SKFGLITELNRSPACHPCSPWCKSGWGESSEDDCSLTRTVCAAGC-ARCKGLPTDCC 236
D 172 PSMTLTVSTIGSSCGRCHKSCG-RCMGPTENHCQTLTRTVCAEQDGRGYPVSDCC 230
QY 237 HEQCAAGCTGPKSHDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASC 296
D 231 HRECAGCGSPKDDCFACMNFNDSCAVTQCPTFYNNPTTFLEHNFNAKYTYGAFCV 290
QY 297 TACPYNLTSDVSGCTLVCPHLNQEVTAEADGTORCEKSKPCARVCYGLGMEHLREVR 356
D 291 KKCPHNIV-VDSSSCVAAACFSSKMEV-ENNGIKCKCKCTDLCIPACDGIIGSLMSAQTV 348
```

```
QY 357 TSANIOEACCKKIFGSLAFLPESFDGDPASNTAPLPEOLQVETLEETIGVYISMP 416
D 349 DSSNIDKFINCTKINGLIFLVYGHDDPNAIDAIDPEKLVNRYTRREITGFNIGTMP 408
QY 417 DSLPLSVFONLOVIRGRRIHNGAYSLTLOGIGISWGLSLRSLGSLALIHNNHLCF 476
D 409 PNMTDFSVFNLSVLTIGRVLVSGSLTLILKQGGITSLQFQSLKEISAGNIYITNSNLCY 468
QY 477 VHTVPMQLRPNRPHQALHTANRPEDCEVGEGLACHOLCARGHGWGPGPOCVNCSQFL 536
D 469 YHTNMWTLFSTVYQRIVIDNRRAENCTAEAGVCMNLCSDNGCWGPGPDCLSCRRFSR 528
QY 537 GOECVEECRYLQGLPREYVNAHCLPCHPECP-ONGSVTCFGEADQCVACAYKDPPE 595
D 529 GKICIESCNLYDEFFRFENGSTICEEDSCGEKMEDELTHGCGPNDCTCSHFQCPN 588
QY 596 CVARCPGVKPDLSYMBIMKPPDEGACOPCPINCTHSVLDLDDKC-----PA 644
D 589 CVEKCPVLOQANSF--IFKADODRECHPCHPCTCGCNGPSTSHDCIYYPWGHSTLPQ 646
QY 645 EGRASPLTISAVN-GILLVVUGVFGILIKRQOKIKRYTMRLLQETELVEPLTPS 703
D 647 HAR-TPL--IAAGVIGLFLVIMALTFAVYVRRKSIK-KKRALRRFL-ETELVEPLTPS 701
QY 704 GAMPNOAMRILKETELRKXVLCGSAFGTYKGIWPDGENVKI PVAIKVLRNTSPKA 763
D 702 GTAPNOQOLILKETELRKXVLCGSAFGTYKGIWPDGETVVI PAIKILNETTPKA 761
QY 764 NKEILDEAYVAVGASPYVSRLLGICLTSTVOLVTQIMPYGLLDHVENRGRIGSODL 823
D 762 NVEEMEDALIMASVDHDLVRLGVCSPITQVLTQMPHCCLLEYVHEHKDNIGSOLL 821
QY 824 NMCMQIAKMSYLEDVVLVNRDLAARVLYKSPHNVKITDQGLARLIDISTEYHAQCGK 883
D 822 NMCMQIAKMSYLEERLVHRDLAARVLYKSPHNVKITDQGLARLIDISTEYHAQCGK 881
QY 884 VPIKMALESILRRPFTHSDVMSGYVTWELMFFGAKPYDGI PAREIPDLLEGERLPQ 943
D 882 MPKMALECIHRKPFTHSDVMSGYVTWELMFFGAKPYDGI PAREIPDLLEGERLPQ 941
QY 944 PPICTIDVYIMVCMWIDSECRPRFELVSEFSRMAADPQRFVIONED-LGPASPLDS 1002
D 942 PPICTIDVYIMVCMWIDSECRPRFELVSEFSRMAADPQRFVIONED-LGPASPLDS 1001
QY 1003 TFRSLLEDMDKGLVVAEELYVPQGFPCDP-----APGA 1039
D 1002 KFFQNLLEDDELDMDMAEEYLP-QAFNIPPIYTSRTRIDSNRSEIGHSPPAYTPMS 1060
QY 1040 GGMVHHRHRSSTPSRSGGDLTLGLEPSEEARPSPLAPSEAGSDVFDGLGMAAKGLQ 1099
D 1061 GSGFVYDDGFPATQGS--MPMTYATTSITLPEAPVA--QCATREMFDDSCNGTLKRPV 1115
QY 1100 SLPTHDSPLQRYSEDPVLPFS-----ETDGYVALTCSPOPEYVNOQVYRPOPSP 1152
D 1116 VPHVQEDSSTQRYASADPTVAPERNPRAIDEBEGVPMHDKPKQEOYLNPVE----- 1167
QY 1153 REGPLPAPARAGATLEAKTSLSPGKNGVADVAFGGAIVENPEYLITQGGAAAPPHPPA 1212
D 1168 -ENPFVSR--KNGDLQ-----ALDNPEYHASSSG-----PPKA 1198
QY 1213 -----FSPAFNLVYWDQDPPEGA--PPT 1236
D 1199 EDEVNEPPLYLNTFTNALGNAEYMKNSLSVPEKAKAFNDPDMVHNSLPRRSTLQHPDY 1258
QY 1237 FKGTPT-----AENPEYL 1249
D 1259 LQESTYKYFYKQNGRIRPIVAENPEYL 1285
RESULT 8
XMRK_XIPMA STANDARD; PRT; 1167 AA.
ID_XMRK_XIPMA
AC_P13388;
```

Query Match	Best Local Similarity	39.7%	Score 2700.5	DB 1	Length 1167
Matches 574	Conservative 165	Mismatches 391	Indels 137	Gaps 27	
4 AALCRGMLLALPRGAAS	----	QVCTGTMTKRLPASRPTHLMRLHYOGCOVGN	59		
8 AALLQ--LLVLAISRCSTDPDRKVCQGSNQM	TM--	LDNHLYLKMKNVSGCNVLE	62		
60 LEIYTLPTNALSFLQDIQEVQGYVLAHNOVROVPLRLRIVRGTLFEDNYALAVLN			119		
63 LEITYTQENQDLFSIOEIVQGYVLAHNEVSTIPLVNLRLIRGQNLIEGNTFLVAMSN			122		
120 GDPNLNTPTVAGSAGRLRELOLRSLEILKGGVLIQRNPOLCYODTILMKDQYIKANSK			179		
123 YQK-NPSS--DYVQVGLKQLDLSNLEILSGYKAVSHNPLLCVETIINMWDIVDKTSNP			179		
180 FIGITELNRSRACHPCSPMCKSGKSCWGSSESDCSLTFTVAGGC-ARCKGRLPTDCCHE			238		
180 TMLNIPHAFEHQCKODHGCYNVSCMAKPRHCKFKTLCLACQCNRCRCKPRKPIDCCNE			239		
239 QCAAGCTGPKHSQCLACLHFNHSGICELHCPALVTYNTDTFESMPNREGRTYFGASCVTA			298		
240 HCAGCGTGPRTDCLACRDFNDGTCQDTCRPPKXIVDIVSHQVVDNPIKXTFGACAYKE			299		
299 CPVNVLTSDVSCITLVGRYLHNOEVLTAEDGTORCEKSKPCARVCYGLAMEHLREVRATVS			358		
300 CPNSVYVTE-GACVRSASAGMLEVD-ENGKRSCKPCQGVCPKVDGIGISLNTIAVNS			357		
359 ANIOEFAQCKKIFGSLAFLPESFGPDASNTAPLOPQLOVFEETLEITGVLYSAMPDS			418		
358 TNIRSFENCTKINDIILNANSFEGDBHYKIGTDPEHLMNLTVKELTGLVLTMMMPEN			417		
419 LPDLISFQNLQVIRGRILHNGAYS-LTLQGLGISMLGRISLRELSGSLAIHNHTHLCFV			477		
418 MTSLSVFQNLIEIRGRITTFSGKGSFVVVUVRHLOMELRSLKEVSAQNVILAKNTLOLRYA			477		
478 HTVWMDLFQNPQALHTANRPREDECVSGSLACHOLCLAHGHCCKRPTQCVNCSQPLRG			537		
478 NTLNWRLLFRSESDISLEYDART-----ENQTCNNKCESDEGCGPRTWCSCIAHIDRG			530		
538 QECVSEECRVAGLIPREYVVARHCLPCHREQORQNGSTCGRRPADQVCANAHYDPRFCV			597		
531 GRVVASGNLQGBREARQVQDGRGVQOQVCLVQDTSILTYGPRPANCSSKANAFQDGPQCI			590		


```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 141 183 EILSGVYIEKPKLCHMOTIDMDIYDRDAELVVDNGR
SC -> GQFVWVSGLTLPQPDWYLLDDDERLLTSSASK
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

```

Query Match 35.7%; Score 2432.5; DB 1; Length 1342;
 Best Local Similarity 40.7%; Pred. No. 3.6e-122;
 Matches 534; Conservative 191; Mismatches 457; Indels 129; Gaps 32;

```

QY GILLALLPFGAA--STOVCTGTMKLRLPASPEPHLDMLRHLYOGCGVQGNLETLVPT 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 GLLFSLARGSEVGNISQAVCPPTLNGLSVTGDAENQYTLKYKERCEVWGNLEIVLGH 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY NASLSFLODIOEVQYVLIANQVRQVPLRLRIVRGTOLFEDNYALAVLDNGDPLNNT 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 NADLSFLQWIREVTGYLVAMNERSTLPENLRVVRGTQYDGFALFV-----LNYNT 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 PYTGASPGGLRELQLRSLEILKGVLIQGNPOLCYODTILMKDQYIKANSKPIGITELN 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 ----NSSHALRQLRLTQLTEILSGVYIEKNDKLCHMDTIDMRDIDVDRDAE---IVYKD 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 188 RSRACHPCSPMCGSRGWSSSEDGSLTRTVCAAGC-ARCKGRLPDDCCHEGCAAGCTG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 NGSSCPCEHCVCG-KCMWPGSSDDCQTLTKTICAPQCGNHGCFGPNPQOCHDEACGCSG 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 PKASDCLACLHFNHSGICELHCPALVYNTDTFESMNPGRGRTYFGASCVTACPYNYLST 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 PQOTDFACRHPFDSGACVRCRQPLVYNLTQLLENNPTTKYQYGGVCAASCPHNFV-V 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 DVGSGCTLVCPHNOEVTAEADGTQRCCKSPKARVCYGLGMEHLREVRAYTSANTIOFAG 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 DQTSRCRACPPDKMEVD-KNGLKMKCEPCGGLCPACGCTGSG--SRQYDSSNIDGFVN 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 CKKIFGSLAFPSFGDPASNTAPLQPELOVFEETLEITGVLYISAWPDSLDLSVFO 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 CKILDNLDPLITGLNGDPHMKPALDPEKLANFRVYREITGYLNTQSWPNNHNFVSFS 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 NLQVINGRIHLNGAVS-LTLQGLGISWLGRLSRLSGLALILHNHTLCFVHTVPDQOL 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 NLTTIGRSLYNGRSLIMKNLNTVSLGRSLKEISAGRIYISANROLCTHHSLNMTKY 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 486 FRNPQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGCEVSEC 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 LRGPTEERLDIKHNRPRRDCVAEGKVCDDLSSGCGWPGGQCLSCSNYSRGCGVTHC 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 545 RVLOGLREYVYNAHCLPCHPECCPONGSVTCFPEPADQCVACAHYDPPRCVARCPSGV 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 NPLNGEPRERPAHEECSCFHECCQPMEGTATCNCSGSDTCAQCAHFRRGPPCVSSCPGV 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 KPDLSYMPIMKFPDEBACOPCPINCTHSCVDLDKCGCPABORA----SPITSIVSAVVG 660
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 594 LG--AKRPITYKPVQNECRCHENCHTQCGKPELDQCLGTLVLIGKTHLTMLTVAAG 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 ILLVVLGVVFGILLIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNOAMRLKETE 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 --LVVIFPMLOGFTLYRGRRIQWKRAMRRYLERGESIEPLDPS-EKANIKVARIKETE 708
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 720 LRKKVUGSAGFVYVGMIPDGENYKIPAIIVYLEBNSPKANKELIDEAIVYMAVGS 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 LRKLKVLGSGVFGTVHKGWIPEGESIKIPVICIYIEDKSGROSFOAVTDHMLAIGSLDH 768
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 780 PYVSRLLGICLSTVQOLVTOLEMPYGCLLDHVRENRGRSGQDLINMCQIAKMSYLEVD 839
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AHVRLGLGCPGSSLOLVTOYLPGLSDHVRQRGALGPOLLNMGVQIAKGMVYLEH 828
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 840 RLYVRDLAARNVLVKSBNHYKITDPGLARLIDIDETEHADGKVPKIMALESTLRERF 899
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 GMVHRNLAAARNVLVKSQVAVDFVADLLPDDQQLVSEAKTPIKMALESIHPEKY 888
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 900 THQSDVMSYGVTTWELMTFGAKPYDGI PAEIPPLLEKGERLPORPCTIDVYIMVYCW 959
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 THQSDVMSYGVTTWELMTFGAEPYAGRLAEVPPLEKGERLAPQICTIDVYIMVYCW 948
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 960 MIISECPRRRELSEFSRMAPDPORFVIONEDLGA---SPDSTFYRSLDEDDMGD 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 MIDENIRPTRELKELANETFRNARDPRYLVIKRS-GEGLARGPPEHGLTNKKLEVELEP 1007
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1017 LVDAEYLVPQGFCDPAPAGAGVYHNRHSSSTRSGGDLTLGLEP-SEEBAPRSP 1075
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1008 ELDDLDLLEABED-----NATLTLSALSLPVGTNLRPGSGSL 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1076 APSEGASDVFDGDLGKAAGLQSLPETHD-PSPLQRYSEDPVPLP-----SETDGYV 1128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1049 SPSGSGY-MPMNOGNTGSCQESAVSGSSERCPRVSLH-----PMRGLCLASSESECHV 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1129 A-----PLGCSPOPE-----YVNPQDVRRQPPRPREP-----L 1157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 TGSEABELQEVSMCRSRSRSPRPRGDSAYHSQHSLLTPVTVLSPGLEEDVNGYVM 1161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1158 PAARPAGATLERAKTISP-GKNGVY-----KDVFAFGAVENPEYLTPOGGAAPORHP 1210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 PDVHLKCTPESREGTSSVGLSSVLGTGEEDD-----EEYEVNRRRRHSP-PHP 1212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1211 PAFSPAFLNLYWD-----QDPERGAPSTFKGTPTAENPEYL 1249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1213 RPSLEELGYEYMDVGSDLASLSGSTQSCPLHPVPIWPTAGTTPDEDEYEM 1263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
ID_ERB3_RAT STANDARD; PRT; 1339 AA.
AC 062799; 062955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase etbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NM_011011;
RN [1]
SEQUENCE FROM N.A.

```

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; Pubmed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.,
 RT "Cloning of the rat ErbB3 CDNA and characterization of the
 recombinant protein.";
 RL Gene 165:279-284 (1995).
 RN (2)
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.J., Koland J.G.,
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; Pubmed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.,
 RT "Expression of neurogins and their putative receptors, ErbB2 and
 ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659 (1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAR.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION OF TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29339; AAC28498.2; -
 DR EMBL: U52530; AAC53050.1; -
 DR HSBP; P11362; IFGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SMO0261; FU_5.
 DR SMART; SMO0219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSSEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 543 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 621 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BFPD1E CRC64;
 Query Match 34.7%; Score 2362.5; DB 1; Length 1339;
 Best Local Similarity 41.0%; Pred. No. 1.9e-118;
 Matches 526; Conservative 172; Mismatches 430; Indels 155; Gaps 35;
 3 LAALCRNGLLALPRGAA---STOVCTGTDMKRLRSPRETHIDMLHLYOGCOVOCN 59
 7 LQVLC---FLSLARGSEMGNSQAVCGTLNGSLSTGVDANQVOTLYKLYEKEVVMGN 62
 60 LELTYLPFNASLSPLODIOEVGYVLAHNOVROPLORLRIVGTQLFEDNYALAVLDN 119
 63 LEIVTGHMADLSFLQWIREVTGYVLAHMFVSFLPBNLAVNGTQVYDQKFAIFVM-- 120
 120 GDPILANTTPVTGASPGGLREQLRSLEITLKGVLIOHNPOLCYQDTTLMKQDYIKANSK 179
 121 ---LNTYNT---NSSHALRQLKFTQLTEILSGVYIEKNDKLCMDITDMD-IVRVGCA 172
 180 FIGTEINRSRACHPCSPMCKSGRCWKESSDQSLRTFVAGGC-ARCKGPLPTDCHE 238
 173 EIVVK--NNGANCPPECHVKG-RCKWGGPDDCQILTKTICAPCCNGRCFPPNPQCCHD 229
 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESHMPNREGRTYFGASCYTA 298
 230 ECAGGCGSPQDTDFACRRFNDSGACVPRCEPLVYNTQLQLEBNPHTKYOGGVCVYAS 289
 299 CPVNYLSTDVSGCTLVCPRLHNOEVTAEEDGTORCEKSKPCARVCYGLMEHLREYRAVTS 358
 290 CPNHFV-VDQFFCYRACPPDMEVD-KHGLMKCEPCGGLCKPACGEGTSG--SRQVDS 345
 359 ANIEFPAGCKKIPGSLAFPPSPGDPASNTAPLQPELOVFEITLLETGYLYISAWDS 418
 346 SNIGFVNCIKILNLDLTLGLNVDPWPKIPALDPELVNFRVREITGYLNTQSWPPH 405
 419 LPDLVSFQNLQVIRGRIILHNGAYS-LTLOGIGISWLGRLSRLESGGLALIHNTHTLCFV 477
 406 MHNVSIVSNLTTIGRSLYNNRGSFLILMKNLNVTSLGRSLKELISAGVVISANQOLCYH 465
 478 HTVPMDOLFENPHQALLHTA-NRPEDCVSGGLACHOLCARGHCKGPGPTQVNCVSQFLR 536
 466 HSLMTRLLRPSSEERLDIKYDRPLGECLEAGKVCDDPLCSSGSGCGPQCLSCRYSR 525
 537 GQECVSGRVLQGPREVVARHCLPGHRECPQNGSVTCGPAPADQVCAHYKDPFC 596
 526 EGVCTVHCNPLQSGPREVHEAQCSPHCECLPMEGISTCNGSSDACARCAHFRIDGHC 585
 597 VARCPGSKBDLSYMPIMKPPDEGACOPCPINCTHSC--VDLDKCGPAPQASPLTIS 654


```

Db      | 586 VNSCHGILG--AKPPIKYTPDAQNECRPCHECTQCGNGPELODCLGAEVLMKSPHLY 643
Oy      | 655 VSAVGIILVVLGVVFGILIKRROOKIR-KYTMRLLOETLEVLPTSPGAMPQOQMR 713
Db      | 644 IAVTVG--LAVLIMLIGSFLVWRGRRIQNKAMRYLGRSGISIEPLDS--EKAKVLYAR 700
Oy      | 714 ILKETELKRVKVLGSGAGFTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKIIDEAVV 773
Db      | 701 IFKETELKRLKVLGSGVGTVMKGIWIPGESIKIPVCIKVEDSGRSQFOAVTDHMLA 760
Oy      | 774 MAGVGSFVYSRLGICLTSTVOLVQLMPEYGLLHVRENRLSODLLNMCOLIAMK 833
Db      | 761 VGSLDHAHIVRLGLCPGSSSLQVLYOYPLGSLDLHVKKHRETLGPOLLNNGVOIAKM 820
Oy      | 834 SYLEDVRLVHRDLAARNVLYKSPNHVKTDFGLARLLDIDETEHADGKVPKIMMALES 893
Db      | 821 YLLEHSMVHRDLARNVWLKSPSOVADFCVADLLPPDDKQLHSEAKETIIMMALES 880
Oy      | 894 ILRRFTHQSDVWSGVYVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYM 953
Db      | 881 IHFCRYTHQSDVWSGVYVWELMTGAEYAGLRALAEIPDLLEKGERLAQPOICTIDVYM 940
Oy      | 954 IMVKCMWIDSECRPRFRELVSFESRMARDPQRFVVIQWEDLGPASPLDSTFRSLLEDD 1013
Db      | 941 VMVKCMWIDENIRPTFKELANFTMARDPYLYIKRAS--GPGTP--PAAPSPVLTKE 997
Oy      | 1014 MGLDVDAEYLVPOQGFPCPPAPGAGGVHRRHSSSTRSGGCDLTGLLEPSEE----- 1068
Db      | 998 L-----QEAELEPEL-----DLDLDLEEEGLATS 1023
Oy      | 1069 -----EAPRSPAPSEGC-----AGSDVPDGLMGAKGLQSLPTH 1105
Db      | 1024 LGSALSLPTGLTRPRGSQSLSPSSGYMPMNQSSLGEXACLDASVILGSEGRSPISLH- 1082
Oy      | 1106 PSLPQRYSEDPVPLPSETDGVY-----APL-----TC-----SPOE-----YNOQDV 1145
Db      | 1083 PIPRGR-----PASSESEGHVTSSELEQKSVCCSRSRSPRPGDSAYHSQHRHS 1135
Oy      | 1146 RPOPPSPREGP-----LPAPAPGATLERAKTLSP--GKNQVY-----KDVAF 1187
Db      | 1136 LLTPVTLPLSPGLEEDNGGYWMPDTHLRGASSSEGLTSSVGLSGTGEDEDD----- 1191
Oy      | 1188 GGAVENPEYLTPOGGAAPQHP 1210
Db      | 1192 ----EEVEYMKRKGSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P0412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EG 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR C910079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RP REVISIONS.

RA Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Litvneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RX ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Bortova D., Botchan M.R., Bouck J., Brockstein P., Broctier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodgson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkask R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RX STRAIN=Daekwanryeong;
RA MEDLINE=85137938; PubMed=2983232;
RA Madsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

```

RT growth factor receptor.";
 RL Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RA MEDLINE=92038942; PubMed=1936959;
 RX Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/rlb alleles: implications for
 RL the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RA MEDLINE=97248481; PubMed=9094709;
 RX Perrimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the Drosophila
 RL EGF receptor.";
 RL Cell 89:13-16(1997).
 CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARMO, WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-MARK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS. CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF ANTIOSEROSA AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF CUTICLE.
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
 CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE PROTEIN.
 CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- TISSUE SPECIFICITY: UNBOUTIOUSLY EXPRESSED IN EMBRYOS, IN LARVAE, UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLGEN OF TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX AND THORACIC AND ABDOMINAL GANGLIA.
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08536.1;
 DR EMBL; AF052753; AAC08535.1; JOINED.
 DR EMBL; K03054; AAAS1462.1; -;
 DR EMBL; K03417; AAAS1460.1; -;
 DR EMBL; K03416; AAAS0965.1; -;
 DR EMBL; K03418; AAAS1461.1; -;
 DR EMBL; AF109077; AAD26132.1; -;
 DR EMBL; AF109078; AAD26132.1; -;
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -;
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -;
 DR EMBL; AF109083; AAD26131.1; -;
 DR EMBL; AF109080; AAD26135.1; -;
 DR EMBL; AE003454; AAF46732.1; -;
 DR EMBL; X02293; CAA26157.1; -;
 DR EMBL; X78920; CAA55523.1; -;
 DR EMBL; X78918; CAA55521.1; -;

DR EMBL; X78919; CAA55522.1; -;
 DR PIR; A00640; GQFE.
 DR HSSP; P13362; IEGK.
 DR FlyBase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGF_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR TrEMBL; P550011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase; Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing; Developmental protein.
 KW
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 Query Match 28.9%; Score 1967; DB 1; Length 1426;
 Best Local Similarity 32.8%; Pred. No. 2; 3e-97;
 Matches 467; Conservative 186; Mismatches 433; Indels 338; Gaps 41;
 24 QVCTGTDMKRLPASPETHLMLRLHYGCGVQVQGNLELYPT-NASLSFLODIOEVOG 82
 100 KICGTGSRRLVSPSKHNNLRNYNCTVQGNLTLPLPNNLDSFLDINREYTG 159
 83 YVLAHNOVROYPLORLRIYVGTOLF-----EDNYALAVLDNGPLNNTPYTGASPGGL 137
 160 YILSHVDVKVKKVPKQIIRGRTLFSLVSEKVALFV-----TYSKM 203
 138 RELQRLSTELKGGVLLQRRPOLCYQDTILMKQOYIANSKFIGITELNSRACHPSP 197
 204 YTLRPPDRDLVNGVGHNNYNNLCHMRTIQSEIVSNGTDAYVYDFTAPRRECPKGE 263
 198 MCKGSRGWESSESDQSLTRTVACAGCA--RCKGPLPTDCGHEQCAAGCTGPKHSDCLAC 255
 264 SCTHG-CWGEGRPNQCKSKYLCSQCAAGRCYGRKRECHLFCAGGCTGPTQDCTIAC 322
 256 LHPNHSICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACPYNYLSTDVSGCTIAC 315
 323 KNFDEAVSKKECPMRKYNPTTYVLETNPGKYAVGATCYVECP--GHLLRDNGACVSC 381
 316 PLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSANIOFAGCKIFGSLA 375
 382 PDKMDKGE-----CVPDNGPCPTKCPGVTLH-----AGNIDSFNCTVVDGNIR 428
 376 FLPSFDS--DPASNTA-----PLQEPQLQVFTLEETITGLYISAPSDSLPDSLVQN 427
 429 ILDTFGSGFDVVANYTWGPRYIPLDPRRREVFSVKSITGLYINIEGHHPGRNLSYRN 488
 428 LQVRGRILNHGAY-SLTQLGLISWGLRSLRLBLSGLALIHNTHLCPHYVFWDLF 486
 489 LETIHGRQLMESMPALAIIVSSLSYSEMRVKKQISSGSVYIOHNRDLCYVSNIRWPAIQ 548
 487 RNPQALLHTANREDECVGEGGLACHQICARGHCGMPPTQCVNCSOPFLRQGECEBGRV 546
 549 KEPQKQVWVWENLADLCERKQITICSDQCNEDCGAGACTDCLCTCKNPNFGTCLADCGY 608
 547 LQGLPREYVNAHRLCPHPECQPNQSGSVTCGPRADQCVACAHYKDPFVARGP----- 601

ID	ERBB_ALV	STANDARD;	PRT;	634 AA.
Db	609	ISNAK- -FDNRCTCKI	CHEPCR- - - - -	TGAGAGDHQCEVNAVDGQHCISECRKAYN 660
Qy	602	-----	SGVK- - - - -	-----
Db	661	DRGVCRECHATCGCTGPGKDTIGIGACTTCNLA	IINNDATVAKLLKDKDCPD- -GY- -FW	-----
Qy	615	KF- -DDEGAACP- - - - -	-----	CP1- - - - -
Db	718	EYVHPOQSLKPLAGRAVCKRCHPLCELTNYGVH	EOVCSKTHYKREOCETECPADH 777	-----
Qy	633	-----	SC- - - - -	VDLDDKG- - - - -
Db	778	YTDDEQRCEQFQHNPECNGCTGPGADCKSCRN	FKFLFDANENGPVYNSTMFTNCKSLPEM 837	-----
Qy	647	R- - - - -	ASPLTS- - - - -	-----
Db	838	RHVNYOYTAIGPYCAASPRSSKITANDVNNI	FIITGAVLVPTICILCV- -TYICRQK 895	-----
Qy	660	OKIRKYT- -MRLLQETELVEPLTPSGAMPNOA	ORIKETELRKYVLYGSGAFCTYVG 737	-----
Db	896	QKAKETKMTALSGCEDSEPLRNSNGANLCKR	IKRYKDAELRGVULGMAFERVYKG 955	-----
Qy	738	IWIPDGENVKIPVAIKVLRENTSPYANKEL	IDEAYVMAGVSPYVSRLLGICLTSTVOLV 797	-----
Db	956	VWVPEGENVKIPVAIKELIKSTGASSEEPLEA	VIYMASSEHVNLKLLAVCMSSQWMLI 1015	-----
Qy	798	TOLMAYGCLLHVHNRNRRLSGODLLNMCMA	AKMSYLEDVRLVHRLAARVLYKSPN 857	-----
Db	1016	TOLMPLGCLLDVNRNRKRGISKALLN	STQIAKMSYLEERLVLHRLAARVLYQTPS 1075	-----
Qy	858	HVKITDFGLARLLDIDETEHYHADGKVP	IKMMALLESILRRRFTHOSDVMSGYVTWELMT 917	-----
Db	1076	LVKITDFGLAKLLSSDSNEYKAAGKMP	IKMLALECINNRVFTSKSDVMATGVTIMELTT 1135	-----
Qy	918	FGAKPYDGIIPAREIPDLLEKGERLP	POPRICTIDVTVMVCKMIMIDSECRPFRELVSEFS 977	-----
Db	1136	FGORHENIIPAKDIPDLIEVGKLEQPEIC	SIDICTLLSCMHLDAARPFKQILTYVFA 1195	-----
Qy	978	RMARPORFVYIONDLG- -PASPLDS	FFYSLIEDD- - -DMGDLVDAEEXLVVQCGFC 1032	-----
Db	1196	EPARDPGYLLAIPGDKFTRLPA- - -	YTSQEKDLIRKLAPTTDSSEAIAPKPDYLQ 1248	-----
Qy	1033	PDPAFAGAGMVHNRHRSSTRS	GGDLTLGLEPSEEAR- - -RSPLASSEGAGSDVF 1086	-----
Db	1249	PKAAPGPS- - - - -	HRTDCT- - - - -	DEMFKLNRKYCKDPSNKSNGSGDER 1287
Qy	1087	DC- - - - -	DICMGAAKGLGSLPTHDPSELQY	SSDPVYPLPSEFSDGYAALPTCSPPQREYVNOP 1143
Db	1288	DSSAREVVGNNLR- - - - -	-----	LDLPVDDDDYLLMP- -TCQPGNNNNM 1325
Qy	1144	DVRPOPPEPRGCPPLAARPA	GATLERAKTILSPGKGVGVKDVAF	ARGAVENPEYL- - - - -TP 1199
Db	1326	N- - - - -	NPNQNNMAAVGVAAGYM- - - - -	DLGVPVSDNPEYLLNAQTL 1364
Qy	1200	CGGAAPORH- - - - -	-----	PPAFSP- -AFDNLTYWD 1224
Db	1365	GVGSESPITPTQIGIPVMGPGTMEV	KVYVMPGPSEPTSDHEYYND 1408	-----
RESULT 12				
AC	ERBB_ALV	STANDARD;	PRT;	634 AA.
Db	P00534;	21-JUL-1986 (Rel. 01, Created)		
Qy	15-JUL-1999 (Rel. 38, Last sequence update)			
Db	15-JUN-2002 (Rel. 41, Last annotation update)			
Qy	DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).			
Db	V-ERBB			
Qy	Avian leukosis virus.			
Db	Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.			
Qy	NCBI_TaxID=11864;			
Db	[1]			

```

RP      SEQUENCE FROM N.A.
RA      MEDLINE=85282822; PubMed=2988784;
RX      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RT      Crittenden L.B., Raines M.A., Kung H.-J.;
RT      "c-erbB activation in ALV-induced erythroleukosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-truncated EGF receptor.";
RL      Cell 41:719-726 (1985).
CC      -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AYIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC      -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/a
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, M10066; AAA48763.1; ALT_INIT.
DR      PIR, A00643; TVCHLV.
DR      PIR, B00643; TVFVLV.
DR      HSSP, P1362; 1FGK.
DR      InterPro, IPR000719; Euk_pkinase.
DR      InterPro, IPR001245; Tyr_pkinase.
DR      Pfam, PF00069; pkinase; 1
DR      PRINTS; PR00109; TYRKINASE.
DR      Prodom, PD000001; Euk_pkinase; 1.
DR      SMART; SM00219; Tyrc; 1.
DR      PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR      PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
DR      Glycoprotein; Phosphorylation.
DR      KW      DOMAIN
DR      FT      NP BIND 132 399
DR      FT      BINDING 165 165 ATP (BY SIMILARITY).
DR      FT      ACT SITE 257 257 BY SIMILARITY.
DR      SQ      SEQUENCE 634 AA; 70891 MW; E705E3JA0B01FCC CRC64;
SQ
Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 3,46-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
QY      587 CAHKKDPFCVCARCPGCVKPDLSWYPIWKPFDEEGACQPCPINCTHSQVLDLDDKCPAQ 646
DB      3 CAHFIDGHCCKAPAGVAGENDTL-VWKYADANAVALCOLCHPNCRTCKGPGLEGCP--- 58
QY      647 RASPLTSVSAVV-GILLVVLGVVPGGILIKRQOKRKRTMKRLLDDELVEPLTSGA 705
DB      59 NGSTPTSIAGVGVGLCLVVGIGIGILVRR-HIVKRTLRLLDERELVEPLTSPGE 117
QY      706 MPNOQMRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVALIKVLRENTSPANK 765
DB      118 APNQAHLRLKETEFKKVKVLGSGAFGVYKGLMIPGEKVKIPVALIKELRENTSPANK 177
QY      766 EILDEAYVMAGVGSPPVSRLLIGICLTSTVOLVTQMLPBYGCLLDHVENRGRLSQDLLNW 825
DB      178 EILDEAYVMASVDPNPHVCRLLGICLTSTVOLITQMLPBYGCLLDVIREHKONIGSYLVNW 237
QY      826 CMOIAKMSVLEDRVLVHRDLAARNVLYKSPNMYKTIIDFGLARLLIDETFYHAADGKVP 885
DB      238 CVQIAKMSVLEDRVLVHRDLAARNVLYKTPHVKITIDFGLAKLLGADKEKYHAEGGKVP 297
QY      886 IKMMALESIIIRRRFTTHOSDWSVGVYTWELMTFCARKYDGIIPAREIPDLLEKESRLLPOPP 945

```

```

Db      1298 IKMMALESILHRIYTHOSDWSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPOPP 357
Qy      946 ICTIDVYMIWVKCMIDSECRPRELVSFESRMAARPORFVYQ-NEDLGPAFLDSTF 1004
Db      358 ICTIDVYMIWVKCMIDADSRPKRELIAEFSKMAARPPRLVYIQGBRNMHLPSTDSKF 417
Qy      1005 YRSLLDDMDGLVDAAEYLVPQGFPCPPAPAGAGVNHRRSSSTRSGGDLTLGLE 1064
Db      418 YRTLMEEDMEDIVDADEYLVPHGFF-----NSPST----- 449
Qy      1065 PSEEBARSP-----APSEGASDVFDGDLGMAAGLQSLPHTDPSPIQRYSEDPTVP 1119
Db      450 -----SRTPLLSLSATSNNSATNCID-----RNGQHPVREDSFYQRYSSDPTGN 495
Qy      1120 LPSET--DGVVAPLTGSPQPEYVNPQVRPQPSRPGRLPAARPAATLERATLSPK 1177
Db      496 FLESIDIDGFL-----PAPEYVNO--LMPKKSTAM-----TMVQ 526
Qy      1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVY 1222
Db      527 NQLYNNISLTAISKLPMDSRQYQNSHTADNPEYL-----NTNQSPLAKTVFESSPY 578
Qy      1223 WDO-----DPE-----RCAPSTFKGTPTAENPEYLGIDVP 1254
Db      579 WISQGNHQLNDPNDYQODFLPNETKPNGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OX Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCBI_TaxID=79685;
OX RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "sequencing the erbB gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06544; TVYUH.
DR HSSP; P11362; IEGK.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146
FT BINDING 165 165
FT ACT_SITE 257 257
FT CONFLICT 140 140
FT CONFLICT 146 146
FT SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

```

Query Match 25.0%; Score 1703; DB 1; Length 604;
 Best Local Similarity 52.2%; Pred. No. 9.6e-84;
 Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

Qy 587 CAHYKDPFCVAPRCSCVKRDLSTYPIWKRPDEGACQPCINCTHSCVDLDDCGPAEQ 646
Db 3 CAHFTDGPCHVCVACPAGVLGENDTL-VRKXADANAVQULCPNCTRCKCGGLGCP-- 58
Qy 647 RASPLTISVAVV-GILVVVLGVFGILIKRROQKIRKYMRLLOETELVEBLTPSGA 705
Db 59 NGSTPTSIAGVGVGGLCLVVGGLGYLRR-HYRKRTIYRLLOERLEVEBLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKYVLSGAGFTYVKGWIIPGSENVKIPVAIKVIRENTSPYAK 765
Db 118 APNQAHRIKETEFKKVYVLSGAGFTYVKGWIIPGSENVKIPVAIKVIRENTSPYAK 177
Qy 766 EILDEAVVMAGVSPVYSLIGTCTSTVOLVQLMYRGCLDHYRENRGLSGODLNM 825
Db 178 EILDEAVVMASVDPHPCRLIGICTSTVOLVQLMYRGCLDHYRENRGLSGOYLNM 237
Qy 826 CMQIAKMSYLEVDRLVHRLDIAARNVLYKSPNHKRTDPSGLARLLDDETYHADGGKVP 885
Db 238 CVOIAKAMNTLEERLVRDLAARNVLYKTPQHKITDIFGLAKLGADKEYHAGGKVP 297
Qy 886 IKMMALESILRRFTTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPOPP 945
Db 298 IKMMALESILHRIYTHOSDWSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPOPP 357
Qy 946 ICTIDVYMIWVKCMIDSECRPRELVSFESRMAARPORFVYQ-NEDLGPAFLDSTF 1004
Db 358 ICTIDVYMIWVKCMIDADSRPKRELIAEFSKMAARPPRLVYIQGBRNMHLPSTDSKF 417
Qy 1005 YRSLLDDMDGLVDAAEYLVPQGFPCPPAPAGAGVNHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDADEYLVPHGFF-----NSPST----- 449
Qy 1065 PSEEBARSP-----APSEGASDVFDGDLGMAAGLQSLPHTDPSPIQRYSEDPTVP 1119
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGQHPVREDSFYQRYSSDPTGN 495
Qy 1120 LPSET--DGVVAPLTGSPQPEYVNPQVRPQPSRPGRLPAARPAATLERATLSPK 1177
Db 496 FLESIDIDGFL-----PAPEYVNO--LMPKKSTAM-----TMVQ 526
Qy 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVY 1222
Db 527 NQLYNNISLTAISKLPMDSRQYQNSHTADNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1219 NLYWDDPPERKAPSTFKGTPTAENPEY 1248

```

Db 575 SSPYWGSGNHQ-----INLDNPDY 594

RESULT 14
ERBB_AVIEU STANDARD; PRT; 540 AA.
ID ERBB_AVIEU
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
ON NCBI_TaxID=103898;
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beng H., Engel J.D.,
RT "A single amino acid substitution in v-erbB confers a thermolabile
phenotype to ts167 avian erythroblastosis virus-transformed erythroid
cells.";
RT Mol. Cell. Biol. 6:1751-1759(1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: M3179; AAA2401.1; -
DR PIR: A25231; TVFVB.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TyrKC.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B5D CRC64;

Query Match 23.9%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 6,5e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 587 CAHYDPDFCVARCPGVPKDSYMPWKPPEDEGACQCPINCTHSCVDLDDKCPAEO 646
Db 3 CAHFIDGPHCVACRPGVGENDTL-VMYADANAVCOCHPCTGCKGPGEGCP--- 58
QY 647 RASPLTSIVSAVV-GILLVVVGVGVGILIKRQKIKRYKTMRLLOETELVEPLTPSGA 705
Db 59 NGSKTPSIAGVVGGLCLVVGGLGLVLRRL-HIVKRTLRLLQERLEVEPLTPSGE 117
QY 706 MENOQAMRLKETELRKVVLGSGAFVYKGIWPDGNNVKAIVAKYLAENTSPKANK 765
Db 118 ANQOHLRLIKETEFKRVVLGFGAFVYKGLMIPEGEKVTIPVAIKELREATSPKANK 177
QY 766 ELIDEAYVAGVSPVSLRLGLTSTVQLTQMLPVCGLLDHVENGRGLSQDLLMW 825
||||| :|:| |

Db 178 ELIDEAYVAGSVDPNPHVCRLLGICLTSTVOLITQMLPVCGLLDYIREHKONIGSQYLLNW 237

QY 826 CMOIAGKMSYLEDVRLVHRDLAARNVLVSPNNHVKITDPGLARLLDIDETFYHADGGKVP 885
Db 238 CVOIAGKMNYLEERHNRHDLAARNVLVSPNNHVKITDPGLARLLDIDETFYHADGGKVP 297
QY 886 IKMMALESILRRRFTHQSDVMSYGVTVWMLTFCAPKPYDGI PAREIDPLEKGERLPQPP 945
Db 298 IKMMALESILHRIYTHQSDVMSYGVTVWMLTFCAPKPYDGI PAREIDPLEKGERLPQPP 357
QY 946 ICTIDVYMTWKCMMDISECRPFRELVSFSPKARPQRFVYQ-NEDLGPAAPLDTSTF 1004
Db 358 ICTIDVYMTWKCMMDISECRPFRELVSFSPKARPQRFVYQ-NEDLGPAAPLDTSTF 417
QY 1005 YRSLLEDMDVDLVADEEYLVPPOGFPCPDPAAGAGMHHRRSSSTRSGGGLTLGLE 1064
Db 418 YRSLLEDMDVDLVADEEYLVPPOGFPCPDPAAGAGMHHRRSSSTRSGGGLTLGLE 449
QY 1065 PSEEEAPRSP-----APSEGASDVFDGLGMAKGLQLPTHDPSPLQRYSEDPVP 1119
Db 450 -----SRTPLSLSLATSNNASATNCIDRNG-----H----- 476
QY 1120 LPSETGYVAPLTCSPOPEYVNPDPVPPSPREGLPAPAPAGAT-LEBAKTLSPCKN 1178
Db 477 -PVREDFGL-----PAPEYVNO-LMPKKPSTAMVQNOIYNYISLTAISKLPIDSRYN 527
QY 1179 GVVKDVPFAGCAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR_CHICK STANDARD; PRT; 703 AA.
ID EGFR_CHICK
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
(Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC Gallus.
ON NCBI_TaxID=9031;
RX [1]
SQ SEQUENCE FROM N.A.
RA MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.,
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
expression in mouse cells, and differential binding of EGF and
transforming growth factor alpha.";
RT Mol. Cell. Biol. 8:1970-1978(1988).
RT FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
dimerization, internalization of the EGF-receptor complex,
induction of the tyrosine kinase activity, stimulation of cell DNA
synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib.ch.

```
CC -----
DR EMBL; M20386; AAA48760.1; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; Fu; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 668 >703 BY SIMILARITY.
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 703
SQ SEQUENCE 703 AA; 77427 MM; AFE2DE11B735A690 CRC64;
```

Query Match 23.6%; Score 1610; DB 1; Length 703;
Best Local Similarity 44.4%; Pred. No. 1e-78;
Matches 314; Conservative 113; Mismatches 252; Indels 28; Gaps 12;

```
QY 8 RMGLLALALPPGAA-----STQVCTGDMKRLRPAPFETHLDMRLHLYOGCCVQVGNLE 61
DB 13 RGAVALVLLGLVALGSAVEEKVCGGTNNKLTQLGHVEDHFTSLQRMVNNCEVVLNLE 72
QY 62 LTYLPTNASLFLQDIQEVGVYLIANQVROVPLQRLIVRGTLQFEDNYALAVLDNGD 121
DB 73 IYVEHNRDLTFLKTIQEVAGVYLIALNWVVIPLNLQIIRGNVLYDNSFALAVLSNYH 132
QY 122 PLANTTPVTCASGGLRELQRLSTELKSGVLIQRNPOLCYODPTIIMKQVYIKANSKFT 181
DB 133 -NAKTO-----GLRELPMKRLSELTLNGVKSNNPKLCNMDVTLAMD-IIDTSRKPL 182
QY 182 GITEI-LNRSBACHPCSPMCKSGSRGCSSEDCOSLITRTVCAGCA-RCKGRLPTDCCHEQ 239
```

```
DB 183 TVLDPASNLSSCPKCHPNCTEDHCGAGEONCQTLTRVITCAQDCSGRCRGVPSDCCHNQ 242
QY 240 CAAGCTGPKASDCLACGIFHNSGICELHCPALVYVNTDTRESMNPREGRYTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACKKPRDDATCKOTCPPLVYINPTTQMDVNPBGKTSFGATCVREC 302
QY 300 PNYVLSTDVGSCTLVCPLIHQEVTAEADGTORCEKCSRPCARVCGYGLMEHLREVAVTSA 359
DB 303 PHNYVVTDHGSCVRSCTVDYEV-EENGVRKCKKCDGLGSRVNCIGIGELKGLISINAT 361
QY 360 NIOEPACCKKI FGSIAFPESPDGPNASNTAPLOPEOLQVFETLEEITGVLYISAMPDSL 419
DB 362 NISFKXCTKINGVSIPLPAFLGDAFTKTLPLDPKKLDVFRYKVEISGFLLIQAMPDNA 421
QY 420 PDISVFNQLQVIRGRIIHNQAVSLTLOGLISWIGRLSRLSELGSLALIHINHTLCVHT 479
DB 422 TDLYAFENLEIIRKRYKHQGYSLAVNVLKIQSLGLSLKEISDGLAIMKNKLCYADT 481
QY 480 VPMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQVCVNSQFLRGQE 539
DB 482 MNMRSULFATQSOKTKIIONNKNKNDCTADRVVCDPLCSDVGCMGPGPHCFSCRFSSQKE 541
QY 540 CVEECRVLOGLPREYVNAHCLPCHPECPQNG---SVTCFGEPADCCVACAHYKDPPEC 596
DB 542 CVXQCNILOGEPRFERSKCLPCHSECLVQNSTAVNTCSGPGPDHCKMCAHFIDGPHC 601
QY 597 VAPCPGVRPDLGYMPIWKFPDEBGAQCPPIINTHSCVDLDKCCPAEQASPLTSTVS 656
DB 602 VKACPAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGPGLEGCP--NGSKTFSIAA 657
QY 657 AVV-GILLVVVLGVFGILIKRROQKIRKXTMRLLQETELVEPLTP 702
DB 658 GVGVGGLCLVVGIGLYLRR-HIVRKRTIARLLQERELVEPLTP 703
```

Search completed: July 22, 2003, 08:44:44
Job time : 20.2304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-149-163-12

Sequence: 1 METALCRMGLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLASTSUM62

Gapop 10.0, Gapex 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriopl:*

17: sp_archaeopl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6225	91.4	1259	6	018735 canis faml
2	3165	46.5	1209	11	090X70 canis norv
3	3133	46.0	1210	11	09EP98 mus musculu
4	2734	40.1	1165	13	09YH40 xiphophorus
5	2705.5	39.7	1137	13	09W6F6 gallus gall
6	2317	34.0	1328	13	P79754 fugu rubrip
7	2032.5	29.8	1433	5	09BIH9 qbblns anopheles g
8	1802.5	26.5	419	4	09UK79 mus sapien
9	1739	25.5	367	11	08R2X1 mus musculu
10	1720	25.3	729	15	086712 avian rous-
11	1718	25.2	567	15	086714 avian rous-
12	1697.5	24.9	412	4	08WYV0 mus sapien
13	1653.5	24.3	962	15	064895 avian eryth
14	1645	24.2	545	15	085468 avian eryth
15	1524.5	22.4	655	11	09WYF5 mus musculu
16	1508.5	22.2	643	11	09ERV6 mus musculu

17	1276	18.7	1193	5	09Y1X8	09Y1X8 ephydacia f
18	1194.5	17.5	1368	5	023821	023821 caenorhabdi
19	1162	17.1	1717	5	026566	026566 schistosoma
20	1141	16.8	527	13	090836	090836 gallus gall
21	1019.5	15.0	478	11	09ESE0	09ESE0 ratius norv
22	956.5	14.0	599	13	09ESH2	09ESH2 gallus gall
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.0	176	11	0923V5	0923V5 ratius norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	08SZW1	08SZW1 drosophila
27	762.5	11.2	1362	13	09PVZ4	09PVZ4 xenopus lae
28	754.5	11.1	311	13	09Y1B2	09Y1B2 xiphophorus
29	737	10.8	1671	11	09NUV5	09NUV5 biophthalari
30	726	10.7	331	4	09BUD7	09BUD7 homo sapien
31	712.5	10.5	1368	13	08UM85	08UM85 paracitichy
32	706	10.4	149	6	09BG66	09BG66 oryctolagus
33	697	10.2	1418	13	093457	093457 scophthalmu
34	689.5	10.1	1369	13	08UM86	08UM86 paracitichy
35	679	10.0	1358	13	073798	073798 xenopus lae
36	663.5	9.7	1472	5	09U5A8	09U5A8 bombyx mori
37	659	9.7	1412	13	08UM84	08UM84 paracitichy
38	648	9.5	1245	13	09YGH8	09YGH8 scophthalmu
39	643.5	9.4	1418	13	08UM83	08UM83 paracitichy
40	639	9.4	1371	11	09QVW4	09QVW4 ratius sp.
41	617.5	9.1	2144	5	09VD94	09VD94 drosophila
42	588.5	8.6	987	11	091YMO	091YMO mus musculu
43	587.5	8.6	1036	4	007912	007912 homo sapien
44	587.5	8.6	1055	11	054967	054967 mus musculu
45	586.5	8.6	1091	4	09UMQ4	09UMQ4 homo sapien

ALIGNMENTS

RESULT 1

018735 PRELIMINARY; PRT; 1259 AA.

ID 018735

AC 018735/

DT 01-JAN-1998 (TREMREL. 05, Created)

DT 01-JAN-1998 (TREMREL. 05, Last sequence update)

DT 01-JUN-2002 (TREMREL. 21, Last annotation update)

DS EDB-2.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT "cDNA cloning of exB-2 from canine mammary gland.";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB008451; BAA23127.1; -.

DR HSP; P11362; IFGK.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_Pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_Pkinase.

DR InterPro; IPR004019; YLP_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF01030; Recep_L_domain; 2.

DR Pfam; PF02757; YLP_2.

DR Prodom; PD000001; Euk_Pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TyKc; 1.

DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 91.4%; Score 6225; DB 6; Length 1259;
 Best Local Similarity 91.1%; Pred. No. 0;
 Matches 1148; Conservative 42; Mismatches 64; Indels 6; Gaps 2;

```

OY 1 MELAALCGWGLLLALLPPGAASVCTGTGMDKLRIPASPEHLDMLRHLVYGCQVYQGNL 60
DB 1 MELAMCMWGLLLALLPSGAAGTCTGTGMDKLRIPASPEHLDMLRHLVYGCQVYQGNL 60
OY 61 ELTYLPTNASLFLDIOEVGVYLIAHNOVQVLOLRIRYRGTOLEPDNALVLVNG 120
DB 61 ELTYLPTNASLFLDIOEVGVYLIAHNOVQVLOLRIRYRGTOLEPDNALVLVNG 120
OY 121 DPLNNTTFTVTSASPGGLRELOLRSLTEILKGGVLIQRPOLCYOITIMKQYIKANSKF 180
DB 121 DPLNNTTFTVTSASPGGLRELOLRSLTEILKGGVLIQRPOLCYOITIMKQYIKANSKF 180
OY 121 DPLNNTTFTVTSASPGGLRELOLRSLTEILKGGVLIQRPOLCYOITIMKQYIKANSKF 180
DB 121 DPLNNTTFTVTSASPGGLRELOLRSLTEILKGGVLIQRPOLCYOITIMKQYIKANSKF 180
OY 181 IGITELNRSRACHPCSPMKSGSRCSGSESDCCSLTRYVCAGGACARCKGRLFTDCCHQC 240
DB 181 LTLIDTNRFSPACPCSPKCKAHCKWGSAGSDCCSLTRYVCAGGACARCKGRLFTDCCHQC 240
OY 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRTPGASCVTACP 300
DB 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRTPGASCVTACP 300
OY 301 YNYLSTDVSGCTLVCPLNHNOEVTAEQTCRCEKSKPCARVCYGLGMEHLREVRATSAN 360
DB 301 YNYLSTDVSGCTLVCPLNHNOEVTAEQTCRCEKSKPCARVCYGLGMEHLREVRATSAN 360
OY 301 YNYLSTDVSGCTLVCPLNHNOEVTAEQTCRCEKSKPCARVCYGLGMEHLREVRATSAN 360
DB 301 YNYLSTDVSGCTLVCPLNHNOEVTAEQTCRCEKSKPCARVCYGLGMEHLREVRATSAN 360
OY 361 IOEFAGCKKIFGSLFLPESFGDPASNTAPLOPOLOVFEFLTEITGLYLSAMPDILP 420
DB 361 IOEFAGCKKIFGSLFLPESFGDPASNTAPLOPOLOVFEFLTEITGLYLSAMPDILP 420
OY 421 DLSVQNTLOVIRGRILHNGAVSLTQGLISMLGRSLRELGSLALHNNHLFCFHTV 480
DB 421 DLSVQNTLOVIRGRILHNGAVSLTQGLISMLGRSLRELGSLALHNNHLFCFHTV 480
OY 481 PMDOLFRNPHOALLHTANRPEDECEVGEGLACHQOLCARHGWGPGPTQCVNCSQFLRGQEC 540
DB 481 PMDOLFRNPHOALLHTANRPEDECEVGEGLACHQOLCARHGWGPGPTQCVNCSQFLRGQEC 540
OY 540 VEECHVLOGLPREYVNAHRCBCHPECOPONSVCFCFBEADQCAACHYKDPPECVARC 600
DB 540 VEECHVLOGLPREYVNAHRCBCHPECOPONSVCFCFBEADQCAACHYKDPPECVARC 600
OY 601 PSGVPRDLSYMPIMKFPDEBAGCOPCPINCTHSCVDLDKGPAPORASPLTSIYSAVVG 660
DB 601 PSGVPRDLSYMPIMKFPDEBAGCOPCPINCTHSCVDLDKGPAPORASPLTSIYSAVVG 660
OY 660 PSGVPRDLSYMPIMKFPDEBAGCOPCPINCTHSCVDLDKGPAPORASPLTSIYSAVVG 660
DB 660 PSGVPRDLSYMPIMKFPDEBAGCOPCPINCTHSCVDLDKGPAPORASPLTSIYSAVVG 660
OY 721 RKVKYLSGAGAGTGYKGIWIPRGENVKTPIVAIKULRENTSPKANEILDEAVVAGVSP 780
DB 721 RKVKYLSGAGAGTGYKGIWIPRGENVKTPIVAIKULRENTSPKANEILDEAVVAGVSP 780
OY 780 YVSRLLGLCLSTVOLVTOAMPYGLDLHVRENRRGLSODLNMCMOIAKMSLYLEBYR 840
DB 780 YVSRLLGLCLSTVOLVTOAMPYGLDLHVRENRRGLSODLNMCMOIAKMSLYLEBYR 840
OY 841 LVHRDLAARNVLKSPNFKITDTEFLARLLDIDETEHADGKVDIKMMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNFKITDTEFLARLLDIDETEHADGKVDIKMMALESILRRFT 900
OY 901 HOSDWSYGVYVWELMTGAKPYDGIIPAREIPDLLEKEKRLPOPICITDVMYVWKCM 960
DB 901 HOSDWSYGVYVWELMTGAKPYDGIIPAREIPDLLEKEKRLPOPICITDVMYVWKCM 960
OY 961 IDSECRPRFRELVEFSSMARDPORFVYIIONEDLPASPDLSTFRSLLEDMDMDLVDA 1020
DB 961 IDSECRPRFRELVEFSSMARDPORFVYIIONEDLPASPDLSTFRSLLEDMDMDLVDA 1020
OY 1021 EEYLVPQGGFCPPDPAQAGGMVHRRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 1080
DB 1021 EEYLVPQGGFCPPDPAQAGGMVHRRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 1080

```

```

DB 1020 EEYLVPQGGFCPPDPAQAGGMVHRRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 1079
OY 1081 AGSDVFDLDMGAAGKIGLSLPTHDPSPLQRYSEPTVPLPSETDGYAPLTCSPQPEVY 1140
DB 1080 AGSDVFDLDMGAAGKIGLSLPTHDPSPLQRYSEPTVPLPSETDGYAPLTCSPQPEVY 1139
OY 1141 NOPDVRPOPSPRESEPLPAAPAGATLER-----AKTISPGKNGVVKVFAFGAVENE 1195
DB 1140 NOPDVRPOPSPRESEPLPAAPAGATLER-----AKTISPGKNGVVKVFAFGAVENE 1199
OY 1196 YLTPDGAAPQPHPPAPAFSDNLVYWDODPPERKAGPSTFKGTPTAENPEYLGLDVPV 1255
DB 1200 YLAPRGAAPQPHPPAPAFSDNLVYWDODPPERKAGPSTFKGTPTAENPEYLGLDVPV 1259

```

RESULT 2

```

OY 909X70 PRELIMINARY; PRT; 1209 AA.
DB 909X70;
OY 01-MAY-2000 (TREMBlrel. 13, Created)
DB 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
OY 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DB 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
OY Epidermal growth factor receptor.
DB Epidermal growth factor receptor.
OY EGFR.
DB EGFR.
OY Rattus norvegicus (Rat).
DB Rattus norvegicus (Rat).
OY Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OY Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
DB Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OY NCBI_TaxID=10116;
DB NCBI_TaxID=10116;
OY SEQUENCE FROM N.A.
DB SEQUENCE FROM N.A.
OY STRAIN=FISHER; TISSUE=LIVER;
DB STRAIN=FISHER; TISSUE=LIVER;
OY MEDLINE=90258888; PubMed=2142466;
DB MEDLINE=90258888; PubMed=2142466;
OY Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
DB Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
OY "A truncated, secreted form of the epidermal growth factor receptor is
DB "A truncated, secreted form of the epidermal growth factor receptor is
OY encoded by an alternatively spliced transcript in normal rat tissue.";
DB encoded by an alternatively spliced transcript in normal rat tissue.";
OY Mol. Cell. Biol. 10:2973-2982(1990).
DB Mol. Cell. Biol. 10:2973-2982(1990).
OY [2]
DB [2]
OY SEQUENCE FROM N.A.
DB SEQUENCE FROM N.A.
OY STRAIN=FISHER; TISSUE=LIVER;
DB STRAIN=FISHER; TISSUE=LIVER;
OY Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
DB Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
OY [3]
DB [3]
OY SEQUENCE FROM N.A.
DB SEQUENCE FROM N.A.
OY STRAIN=FISHER; TISSUE=LIVER;
DB STRAIN=FISHER; TISSUE=LIVER;
OY Guttridge K.; Dawson T.L.; Earp H.S.;
DB Guttridge K.; Dawson T.L.; Earp H.S.;
OY Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DB Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
OY EMBL; M37394; AF14008.1; -
DB EMBL; M37394; AF14008.1; -
OY HSSP; P11362; IFCK.
DB HSSP; P11362; IFCK.
OY InterPro; IPR000494; EGFR_L domain.
DB InterPro; IPR000494; EGFR_L domain.
OY InterPro; IPR000719; Euk_pkinase.
DB InterPro; IPR000719; Euk_pkinase.
OY InterPro; IPR002174; Furin-like.
DB InterPro; IPR002174; Furin-like.
OY InterPro; IPR001245; Tyr_pkinase.
DB InterPro; IPR001245; Tyr_pkinase.
OY Pfam; PF00757; Furin-like; 1.
DB Pfam; PF00757; Furin-like; 1.
OY Pfam; PF00069; pkinase; 1.
DB Pfam; PF00069; pkinase; 1.
OY Pfam; PF01030; Recep_L domain; 2.
DB Pfam; PF01030; Recep_L domain; 2.
OY PRINTS; PR00109; TYRKINASE.
DB PRINTS; PR00109; TYRKINASE.
OY ProDom; PD000001; Euk_pkinase; 1.
DB ProDom; PD000001; Euk_pkinase; 1.
OY SMART; SM00261; FU; 3.
DB SMART; SM00261; FU; 3.
OY SMART; SM00219; Tyrc; 1.
DB SMART; SM00219; Tyrc; 1.
OY PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DB PROSITE; PS00107; PROTEIN KINASE ATP; 1.
OY PROSITE; PS00103; PROTEIN KINASE DOM; 1.
DB PROSITE; PS00103; PROTEIN KINASE DOM; 1.
OY KATP-binding; Receptor; Tyrosine; Tyrosine-protein kinase.
DB KATP-binding; Receptor; Tyrosine; Tyrosine-protein kinase.
OY SEQUENCE 1209 AA; 134891 MW; 96FE57F6CC1B773 CRC64;
DB SEQUENCE 1209 AA; 134891 MW; 96FE57F6CC1B773 CRC64;

```

Query Match 46.5%; Score 3165; DB 11; Length 1209;
 Best Local Similarity 50.0%; Pred. No. 6.4e-230;
 Matches 640; Conservative 166; Mismatches 335; Indels 118; Gaps 26;

```

OY 3 LAALCGWGLLLALLPPGA-ASTOVCTGTGMDKLRIPASPEHLDMLRHLVYGCQVYQGNL 61
DB 3 LAALCGWGLLLALLPPGA-ASTOVCTGTGMDKLRIPASPEHLDMLRHLVYGCQVYQGNL 61

```

D	b	15	LAALCAG-----GALBKKVVOGGSINRLTOLGTFEDHLSIQRPNNCEVVLGNLE	66
Q	y	62	LTYLPTNASISFLQDIOEVQGYVLI,IAHNOVRQVPLQRLIRVNGTOLFEDYVALVLDNGD	121
D	b	67	ITYVQRNVDLSFLKTIQEVAGVLI,ALMTVERIPLBNQIIRGNALYENTYALAVLSN--	124
Q	y	122	PLANTTPVTGASPGGLRELOJLSLTLEILKGYL,IOBNPOLCYQDTITLMKDOYIKANSKFI	181
D	b	125	-----YGTNKQTLRELPMNLOEILIGAVRFSPNNPILCMETIORMD---IVQDFVL	173
Q	y	182	GITELNRSR---ACHRPGCMKSGNRCMWESSSDQSLRTRVYAGCA--RCKGRPLTDDCH	237
D	b	174	SMSMDVORHLTGCKCDSPSCNGSCMGWGBENCCOKLTIKICAOCCSRRCRSPSDCH	233
Q	y	238	EQACAGCTGPKASDCLACLFHNSGISCELHCALVTYNTDTEFSMPNEGRTYFASCVT	297
D	b	234	NQACAGCTGPRESDCLVCHRFDEATCKDQTCERPLMLYNPTTYQMDVNBGAKTSFATCVK	293
Q	y	298	ACPYNVLTSDVGSCTLVCPRLHNOEYTABDGTORCEKSKPCARVCYGLAMEHREAVT	357
D	b	294	KCPRYVVTDBHGSCTYRACGPDYEV--EBDVYSKCKCKCGPCRCVKNCGIGIGEPKOTLISN	352
Q	y	358	SANTOEFGCKKIFESLAFLESFPGDPAASNAP,OPOLQVFEFLTEITGLYISAPD	417
D	b	353	ATNKHFFYCTAISDHLIPLPAFGDSEPTRP,LPDRELEITLKTKVETLGEFLIQAMPE	412
Q	y	418	SLPDLISVQNTQVIRGRI,ILHNGASLT,LOGLGISW,IGLSRLSELSSGAL,IIHNTHLCFV	477
D	b	413	NMTDLHAENLEIRGRTRKQHQFSLAVVGLNTISGLRS,LSKEISDGVIIISGNRLCYA	472
Q	y	478	HTVPMDOLFRNPHQALHTANRPEDECVGEGSLACHQLCARGHCMWGPPTQCVNCSQPLRG	537
D	b	473	NTIIMWKLFEPNQKTKIMNRAEKDCKATNHCNPLCSECGWGPBPTDVCSCQNSRG	532
Q	y	538	QECVEECGNVLOGLPREYVNAHCLPCHRECOQONSVCFCGEADQCAACAHYKPPRCV	597
D	b	533	RECVDCNKILGEPEFEFENSECICQCHBCLPQTNWITICTGGRPNCKICAHYVDPHCV	592
Q	y	598	ARCSGVPRLDSYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCGPAEORASP--LTSIVS	656
D	b	593	KTCBSGIMGENNTL--VMKFPADANNVCHLCNANTYCGCAGPLKGC--QOPBSPKIPSTAT	649
Q	y	657	AVVGLILVVVLGVVFGI--LIRKROQIRKYTRMRLLOETELVEPLTPSGAMENOQMRIL	715
D	b	650	GIVGGLFLIV--VALGIGLFMRROQVLRKTRLRLLQRELVEPLTPSGEABNQHLIL	708
Q	y	716	KETELARKVUGSAGFTVYKGIW,IPDEBNKXIPAIATVLRNTSPKANKELIDEAAYMA	775
D	b	709	KETERKKIKVLSGSGFTVYKGLW,IPBEKXKIPAIKELREATSPKANKELIDEAAYMA	768
Q	y	776	GVGSPYVRLIGICLTSIVOLVTO,MPGCLLDHRENGRIGSGDILLNMCQIAKGSY	835
D	b	769	SVDPNPHVRLIGICLTSIVOLITQ,MPYGCCLDYREHNDNGSYLLNMCQIAKGNV	828
Q	y	836	LEDVRLVHRDLAARNVLKSPNHVKITDPGLARL,LDIDETEHADGKVP,IKMALSESIL	895
D	b	829	LEDRLRLVHRDLAARNVLKTPQHVKITDPGLAKLIGAEKEYNHAGKVP,IKMALSESIL	888
Q	y	896	RRRFTHQSDWSYGTTWELMTFFGAKP,PDGIPARIPDLLEKGEFLPOPPICTIDVYIM	955
D	b	889	HRITTHOSDWSYGTTWELMTFGSKP,PDGIPASEIISILEKGEFLPOPPICTIDVYIM	948
Q	y	956	VKCMNIDECRPRFRELSEFSRMA,RPORFV,IO--NBDLGPASPLDSTFRSILEDDM	1014
D	b	949	VKCMNIDDSPPKRELILBFSK,AKARDQRYLIV,IOGDRMHLPSFTDINFYRLAMEEDM	1008
Q	y	1015	GDLYDAEBEYLVPOQGFPCPDPAPGAGVYHHRHSSTRSGGDTLGLLESSEBEPSP	1074
D	b	1009	EDVYDADEYLV,POQGF-----NSPST-----SRT	1034
Q	y	1075	LAPBEGAGSVFDDGLGMAKGLQSLPTHPSPLQRYSEDPYVLPBPT--GGYVADLT	1132
D	b	1035	LLSLSIASNSN-----SSTVACINRNGSCVYKDAF,IOFRTYSPTSVLT,TDNIDITFL--	1086

```

QY      1133 CSPOPEVYNQDVARPOPPSPREGEGLPAARPAAGATLEAKTLSPKNGVWADVFAGGAVE 1192
Db      1087 --PPEPEYINQ--SVPKRPAGSVQVNVYHNPQPLHP-----APGRDLHYQN--PHSNAVS 1133
QY      1193 NPEVLT-TPQGGAAQPHPPAPFASPAFNLYWMD-----DP-----PERGAPPS 1235
Db      1134 NPEVLTAAQ-----PTCLSSGFFDSSALMTOKSHQSHSLNDPNFYQODFFPKAKPNG 1184
QY      1236 TFKGTPTAENPEVLTGLDVP 1254
Db      1185 IFKG-PTAENNAEYLRVAVP 1202

RESULT 3
Q9EP98      PRELIMINARY;      PRT; 1210 AA.
AC      Q9EP98;
ID      01-MAR-2001 (TREMBlrel. 16, Created)
DT      01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT      01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE      Epidermal growth factor receptor isoform 1.
GN      Egfr.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C3H/101, 129/SVJ, AND 129/SVEVATAC;
RA      Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Daniels A.J.,
RA      Sinclair C.S., Fearall R.S., Green P.J., Yee D., Lampland A.L.,
RA      Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA      Mañlle N.J.;
RT      "Comparative genomic sequence analysis and isolation of human and
RT      mouse alternative Egfr transcripts encoding truncated receptor
RT      isoforms.";
RL      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J;
RA      Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Daniels A.J.,
RA      Schehl C., Fearall R.S., Green P.J., Yee D., Lampland A.L.,
RA      Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA      Mañlle N.J.;
RT      "Comparative genomic sequence analysis and isolation of human and
RT      mouse alternative Egfr transcripts encoding truncated receptor
RT      isoforms.";
RL      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF275366; AACG28045.1; -.
DR      EMBL; AF275364; AACG28045.1; JOINED.
DR      EMBL; AF275365; AACG28045.1; JOINED.
DR      EMBL; AF275367; AACG24386.1; -.
DR      HSSP; P13362; 1FGK.
DR      MGD; MGI:95294; Egfr.
DR      InterPro; IPR000345; CytC_heme_bind.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR002290; Ser_thr_pkinase.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00261; Fu_5.
DR      SMART; SM00220; S_TKc; 1.
DR      SMART; SM00219; TYTKc; 1.
DR      PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR      PROSITE; PS00107; PROTEIN_KINASE_AFP; 1.
DR      PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```


SQ SEQUENCE 1165 AA; 129614 MW; 7F7EB380871A7AB CRC64;
 Query Match 40.1%; Score 2734; DB 13; Length 1165;
 Best Local Similarity 45.7%; Pred. No. 2,2e-197;
 Matches 581; Conservative 165; Mismatches 386; Indels 140; Gaps 29;

QY 1 MELALCRWGLLALLPG-AAST---QVCTGDMKRLPASPETHLMDLRLHLYGCGV 55
 DB 4 LELLEL---LLLLLISIGRCCSTDPDRVCCTSNQMTM---LDNHVYLMKMKMGSGCNV 56
 QY 56 VQGNLELYTLPTNASLFLQDIQEVGYVLIHNOVROYPLQRLIRVGRTOLEFEDNYALA 115
 DB 57 VLENLEITTOENQDLSFLOSIOEVGYVLIHNEVSTIPLVNLRLIRGNLVEGNFTLL 116
 QY 116 VLDNGDPLNNTTPTVGASPGRLRELQRLSLTEILKSGVLIQRPOLCYDITLMDQYIK 175
 DB 117 VMSNVOK-NPSSP--DVGVLKQQLQSLNLTLSGAVVSNHNPILLCNVETIMMMDIVK 173
 QY 176 ANSKRTIGITELNRSRACHPCSPMCKGSRGWSESSDQSLTRTVCAAGC-ARCKGPLPTD 234
 DB 174 TSNPTMNLIPHAFEROCCKDPCGVGSCMAPGPHGQKFTKLLCAEQCNRRRCRGKPID 233
 QY 235 CCECCAGCTGPKSDCLACLFHNSGICELHCPALVYNTDTFESMNPBEGRYFGAS 294
 DB 234 CCEHCAGGCTPRATDCLACRDFNDGTCOTCPPKLYDIVSHQVNDPNIKYTFGA 293
 QY 295 CTTACPYNYLSTDVGSCTLVCPLNQEVTAEDGTQRCXKRCARCAVGLGMEHLREVR 354
 DB 294 CKECPNSNVYTE-GACVRSAGMLEVD-ENGKRCRCPDGVCPVCCGIGIGISLNTI 351
 QY 355 AVTSANIOEFACCKIFGSLAFLPESFDGPPASNTAPLOPEOLQVETLEITGYIYA 414
 DB 352 AVNSTNIGSFNSCTKINGDIILNRNSFEQDPHYKIGPMPEHLMNLTVEITGYIYIMW 411
 QY 415 WPDLSLPSVFNLOVIRGRILHNGVYS-LTLOGLISWGLRSBELSGLALIHNTH 473
 DB 412 WPEINTSISVFONLEITIRKRTFSRGFSVVOVSHLQWLGSSLSLEVAGAVILKNTPO 471
 QY 474 LCFVATVPDQLFRNPQALHTANRPEDECVGEGLACHQLCARGHCMGPGFTQCVNCSQ 533
 DB 472 LRYASTIMMRRLFRESDQSI EVDART-----ENQTCNNECSBDCMKGPTMCVSCIAH 524
 QY 534 FLRGECEVBEQVNLQGLPREYVNAHCLFCHPECCOPQNSVTCFGEADQVACAHYKDP 593
 DB 525 VDRGRCAVASCNLLQGEPRQAVDGRCVQCHOECIVQTSITLCYCGPANCSCAKAFQDG 584
 QY 594 PCVVARCSGVKPDLSYMPIMKFPDEBEGCOPCINCHTSQVLDLCKGPAQRASPLTS 653
 DB 585 POCIRCPHGMIGDGDTL-IMKRYADMGQCPCHONCTQGCSPGLSGCRGD-IVSHSSL 642
 QY 654 IVSAVAGLILVVLGVFGLIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMR 713
 DB 643 AVGLVSGGLITVIALLVILVLRRIK-RKRTIRLLOEKLVLELTSGAPNOAFIR 701
 QY 714 ILKETELKRVKVLGSGAGSTYKGIWIDGENVKIPVAIKVIRENTSPRANKELIDEAVY 773
 DB 702 ILKETEFKDRYLGAGAFGTIVKGLMNPGENIRIPVAIKVIRENTSPRANKELIDEAVY 761
 QY 774 MAGVCSPVYSRLIGTCTSTQVLTQVLMFYGLLVHVENRRLSODLLNCMQIAKAM 833
 DB 762 MASVDHPVCRILGICLTSAVOLVQVLMFYGLLVYRHOIRICGQVLLNLCVOIAKAM 821
 QY 834 SYLEEDVRLVHDLAARNVLYKSPNHKITDPGLIARLLDIDETEHADGKVPKIMMALES 893
 DB 822 NYLEBRHLVHRDLAARNVLYKSPNHKITDPGLIARLLDIDETEHADGKVPKIMMALES 881
 QY 894 ILRRFTQSDVWSVGVVWELMTFGAKPYDGI PARETIDLEKEKRLPQPICTIDVYM 953
 DB 882 ILQWITYHQSDVWSVGVVWELMTFGSKPYDGI PAKETIASVLENERLPPQPICTIEVYM 941
 QY 954 IMVSCMIDSECRPFRELVSEFSRMARDPQRFVIVIONEDLGPASPLSTFRSLLEDD 1013
 DB 942 IILKCMIDPSSRRPFRELVSEFSQMAADPSRIVYIQG---NLPSPSDRRLPSRLSSDD 998

QY 1014 MGDVDAEYLVPOQGFCCPDPAAGWVHRRSSSTRSGGDLTLGLEPSEEAPRS 1073
 DB 999 --DVVDADEYLV-----RYKRIIN-RQGS-----E 1019
 QY 1074 PLAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPLQRYSEDPTV-PLPSETDGYAPLT 1132
 DB 1020 PCIPPPNGH-----PVRENSIALRYISDPTQNALPEKDLOH----- 1054
 QY 1133 CSFPQEVNQPDRPOP-----PSPRE-----GPLP-AARPAATLEPAKTLSPCKNG 1179
 DB 1055 -----BYVNOPSETSSRSLSDIYNVYEDLTDGCMGPVLSLSQAEITNFSREYLVNTQNS 1109
 QY 1180 VWKDVAFGAAVENPEYLTPOGGAAPQHPHPAFSPAFLVYWDODPPERGAAPSTFFKG 1239
 DB 1110 L---PLVSSGSMDDPPY---QAG-----YQAAF-----LPOTGALTGNGMF 1144
 QY 1240 TPTAENPEYIGL 1251
 DB 1145 LPAENLELYIGL 1156

RESULT 5
 Q9W6F6 ID Q9W6F6 PRELIMINARY; PRT; 1137 AA.
 AC Q9W6F6;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERBB4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCB1_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 RT embryonic chick hindbrain".
 RL Mol. Cell. Neurosci. 13:237-256(1999).
 DR EMBL; AF121963; AAD31764.1; -.
 DR HSSP; P11362; IFCK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001368; TNFR_C6.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 1.
 DR Pfam; PF02757; YLP; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWNM_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 FT 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 39.7%; Score 2705.5; DB 13; Length 1137;
 Best Local Similarity 47.0%; Pred. No. 3e-195;
 Matches 535; Conservative 170; Mismatches 357; Indels 77; Gaps 24;

QY 161 LCYQPTILMKQYIKANSKFIGITELNRSRACHPCSPMCKGSRGWSESSDQSLTRTV 220

```

Db      3 LCFADT1HMOD1VRNPMWASFTLVPITNGSSGCGRCHKSCGTG-RCMGTEPHNCQTLLTKTVC 61
Qy      221 AGGC-ARCKGLPTDCCBOCAAGCTGPKKSDCLACHPHNSGICELHCPALVTYNTDTF 279
Db      62 AEQDGDGCTGPPYSDCCCHREAGCGSPKQDTDCACNMFDSGACTGQCPTEFYANPTTF 121
Qy      280 ESMNPBEGRYTFGASCVTACPYNYLSTDVSGCTLVCPHLHNOEVAEDGTORCEKSKRCA 339
Db      122 QLEHNNAKATYTGAFVCYKCKPRHNFV-VDSSSCVACACSSKMEV-EENIGIMCKRCTD1CP 179
Qy      340 RVCYGLAMEHLREVRVATSNANIOEFACCKIFGSLAFIPESFDGDPASNTAPLOPELOV 399
Db      180 KACDGICTGSLVAQIVDSSNIDFICTKINGMLIFLVGIGHDPHTHTAIIINPEKINI 239
Qy      400 FETLEETGYLYISAMPDSL.PDLVSFQNLQVIRGRIIHNAGVSLTLOGIGISWGLSLR 459
Db      240 FQTVREITGYLNTQSWPENNTDFRVSNTLVTIGRALYSGLSLILKQGGTSLQFQSLK 299
Qy      460 ELGSGGLALIHNTLHNTLCPVHTVPMQDLFRNPHQALLHTANPEDECEVSGLAGHQLCARGH 519
Db      300 QISAGNIYITDNSLVCYHTYVNTSLFSTPSQKTVIHRNKKAENCTADGVCNBLGSSDG 359
Qy      520 CWGPGPTQCVNCSQFLRGQCEVCRYLQGLPREYVNAHCLPCHPCOP-QNGSVTCFG 578
Db      360 CWGPGPQCLSCRFIRGTCIESCNLYDEFFREFAGSVCMEDPCQCKMEDMNTICYG 419
Qy      579 PEADQVACAHYDOPFCVACRSPGVK.PDLVSYMPIMKFPDEGACQCPICINCHSCVDLD 638
Db      420 PGPDHCKCFHFDGPNVCVEKCPGLOGANSF--IFKXADERCHPCHPCTGCCGPA 477
Qy      639 DKGC-----PAORASPLTISVAVV-GILLVVVLGVVFGILIKRQCKIRYT 686
Db      478 SHDCIYVPMTRQSTLPQHAR-TPL-IAAGVIGGLFIIVIMGLFAVYVRKSKIK-KKRA 533
Qy      687 MRLLQSTELVEPLTBSGAMPNOQOMILKETEIRKRYVUSGAFGYTYKIM.PDGBNV 746
Db      534 LRRL-ETELVEPLTBSGTAPOQOLRIKETEIRKRYVUSGAFGYTYKIM.PDGBNV 592
Qy      747 KIPALIVLBENSPKANKELIDEAVMAGVSPYBRILGICTSTVOLTOIMPGL 806
Db      593 KIPALIKILNETTGPKANVEFMDALIMASMDHPLVRLGVCSPITQLTQIMPGL 652
Qy      807 LDHVRNRLGSGODLLNMCQIAKGSYLEDELVRHDLAARNVLKSPNHVKITPGL 866
Db      653 LDYVHEKDNIGSOLLNMCVQIAKGMVYLEERLVRDLAARNVLKSPNHVKITPGL 712
Qy      867 ARLLDIDETEHADGCVPIKMMALLESILRRFTHSDVMSYGVYTWELMTFGAKPYDGI 926
Db      713 ARLLGEDEKYNADGKMPKMMAL.ECIIHYRKFTHSDVMSYGVYTWELMTFGAKPYDGI 772
Qy      927 PAREIPDLLEKGERLPQPICTIVYIMVCMIDSECRPRPRELSEFSRMARDQRF 986
Db      773 PTEIIPDLLEKGERLPQPICTIVYIMVCMIDSECRPRPRELSEFSRMARDQRF 832
Qy      987 VVIQNEB-LGPASPLDSTFYRSLLEDDMDGLVDABEYLVPOQGFPCPDAPAGAGVNH 1045
Db      833 LVIQGDDRMKLPSPNDSKFFQNLDEEDLMDMDABEYLV-QAFTNIPPIYLSRTIDS 891
Qy      1046 RHSSSSRSGGGLTTLGLESEEARPS--PLAF-SEGASDVVDGDLGMAAGLOSLP 1102
Db      892 NRNOFYVRDGGVYAAEGGV-PMYRPAFCIIPEAVAGATATAEIFEDCCNGTJLRKQVATL 950
Qy      1103 THDPSPLOKSYSEPTVLPSP-----ETDGYVALTCSPOPEVNVQPDVAPQPPREG 1155
Db      951 AKEDSSQRYISADPTVIFPERVIRGELDEDCQYMTPMDKPTDILNVEENPFYSRKNQ 1010
Qy      1156 PLPA-PPAGATLERAKTSLPGKGVKDYF-----AFGAIVENPEYLTPOGAAPQ 1206
Db      1011 DLQAVNDPEYHN-----APNGQPKADEYVNEPLYNFTANTLENAEYL-----K 1055
Qy      1207 PHPPAPSPAFDNLVYWDQDPPERGA--PPSTFKGTPT-----AENPEYL 1249

```

```

Db      1056 NNLEPKAKKAFNDPDYNNHSLPPRSTLQHPDYLQESTKYFYKONGRIRPIVAENPEYL 1114
RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE EtdB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC NCBI_TaxID=31033;
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; Pubmed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 146 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Euk_Dkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_Dkinase; 1.
DR SMART; SM00261; Fuf_3.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KM ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 34.0%; Score 2317; DB 13; Length 1328;
Best Local Similarity 40.9%; Pred. No. 8,2e-166;
Matches 526; Conservative 153; Mismatches 417; Indels 190; Gaps 32;

Qy      9 WGLLALLPP--GAATQ-----VCTGDMKLRLPASBETHLMDLRHLVQCGVQVQNIEL 62
Db      4 WRLLMCVASRLRAASSQTOEAVCPGTONGLSSTGSDENYNNLKKDRYKGEIIMGNLEI 63
Qy      63 TYLPNTASLSFLQDIOEVGVYLIANQVROPFLQRLIRVGTQLFEDNVALAVLDNDP 122
Db      64 TQIESMNDPFLKTRIREVTGVLIAHMHFGEIRPLGQLRVIRGNSLYERRALSVFLN--- 120
Qy      123 LNNTPVTGASPGGLRELQRLSLEILKGVLIQRNPQLCYQDTILMKDQYIKANSKFIG 182
Db      121 ----YPRDG--PSGLNQLGMLNLEILIDGVQIINNRYLRGVPVYWRD-IIRNNDAPIE 173
Qy      183 ITELNRRACHPSPKPKSGRCGSESESDQSLTRTYCAGGC-ARCKGLPTDCCHEQA 241
Db      174 I-QFNGRGV--CHKSC-GNYCWPGRQCOOILTKTYCACPQCDNRGCTSPRDCHECA 229
Qy      242 AGCGPGRGSDCLACIHNHNSGICELHCPALVTYNTDFFESMPNBEGRYTFGASCVTACPY 301
Db      230 AGCKGRLDTCFCACRLFNDSGACVCPQCPQLIYNKQTFQMETNPNATKYQIGSIVSQCP 289
Qy      302 NYLSTDVSGCTLVCPHLHNOEVTAEQTOR-CEKSKPCARVCYGLGMEHLREVRVATSAN 360
Db      290 HFV-VDSSSCVSVCPDKMEV--ERGSQRCCLSGCLCPKVCBGTGAE---QRTVDSSN 343
Qy      361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFEETLEITGYLYISWPSLP 420
Db      344 IDSFINCTKIQGSILHFLVLTGIDDDFKNVPLDAAKLEVERTVAEITDILNIOGWPKELN 403

```

```

OY /421 DLAVFONLQVIRGRILHNGAVSLTLOGISWLGRLSRLREGSLALIHNTHLCFVHTV 480
DB 404 DLSVFSLLTIIQGLSLFRSRLMWRIPTLTSLSRLSRLSDSVISQNAHLCYHNTV 463
OY 481 PMDLFRNPH-QALLHTANREDECVEGEGLAGHOLCARGHMGPGPTQCVACSGOFLRQOE 539
DB 464 NMQLFPGSRVANSLSNRMACVADGRVCDPLCSGSGMGPGPOCLSCRYSRGT 523
OY 540 CVEECRLVQLPREYVNAH-CLPCHPECPONGSVTCFGEADQCVACAHYKDPFECVA 598
DB 524 CVAGCHFNCSGIPREFAGLNGVCVACHPECKPQTGKASCTGGADECMACIKFRDGPYCMS 583
OY 599 RCPSGVPRDLSSYMIKPRDEEGACQCPRLNCHSCVDLDDKCGPAERASPLISYSAV 658
DB 584 SCFAGVA-DEKGLIFKFPNREGCEPHONCTGCGSGPGLNDC---LEARLITSSGOI 639
OY 659 VGLLVVLGVVF-----GILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAQ 711
DB 640 TGIALGVPAGLIFELVLVFLGLMLVHRGLAIRKRAMRYLESGLSFEFLG-GEKTKVH 698
OY 712 MRLIKETELRKVKVLSGAGFGCTVKGWIPIDGENVKIPVAIKVLRNTSPRANKELIDEA 771
DB 699 ARILKPEDLKRKIKPLGSGVFGVTSKGFWIPGEGTVKIPVAIKTIQDSSGRQTFEITDHL 758
OY 772 YVAVGVSFVYSRLGI CLTSTVQLVTLQMLPYGCLLDHVRNRRGLSODLLNMQIAK 831
DB 759 LMSGSLDHPYIVRLGICPGICQLVTLQSSHSLLEIRHKSLDPLQRLNMCVOIAK 818
OY 832 GMSYLEDVLRVHRDLAARNLVKSPNHVKITDFGLARLLDDETEYHADGKVPKIMMAL 891
DB 819 GMYLLEHRVVKHKLARNLILKNDYQVQISDYGVADLLYDDKIVYSEKTKIKMAL 878
OY 892 ESTLRRRFTQSDVWSYGVTVWELMTGAKPYDGI PARERIPDLLEKGERLPEPDICTYD 951
DB 879 ESTLFRYTHQSDVWSYGVTVWEMMSFGAEFYASVOQEVSVLEKEERLSQPAICTIDV 938
OY 952 YMIWVKCMMDISECRPRELVSFSESMARDPORFVITQONEDLGPAPLSTFRSLLED 1011
DB 939 YMWVKCMMDENIRPFKELASDFTMARDPRIYLVIRMEG-----ED 982
OY 1012 DMGGLDVADEEYLVPOQGFPCPPAPAGAGVHHHRSSSTRSGGDLTLGLPSEEAR 1071
DB 983 SGNGEFL-----RGRSER---GILEADLEDEEE-- 1008
OY 1072 RSLPLASSEGAGSDVFDGLMG---AAKGLSLPTHDPSPLO-----RYSEDPTV 1118
DB 1009 -----GLGDRFATPSLQPSPSMSTSPQINSYMWMTQLRYD----- 1044
OY 1119 PLPSETGVYVAPLTCSPQ- EYVNO-----PDVRRQPSREGPL--PAAR 1161
DB 1045 -FVSGGSHIGYLPMSFSPVDITIRQLWYORSRLSVRTLPRSAFRSSREAELEDGAQ 1103
OY 1162 PACATLERAKTILSPKNGVAVKDVAFGCAVENPEYLLPQCGAARQPHAPPAFSPAFNLY 1221
DB 1104 CAGIFPRV-----FSGERN-----PQGG----- 1122
OY 1222 YMDQPEPERGAPSTFKGPTAENPE 1247
DB 1123 --QQRKLSTASSPSFSFKTMADEEDE 1146

```

```

OC Anopheles.
OX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUA.
RA Lycett G.O.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NCV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSSP; P11362; JFGC
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; EGFR_L_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fu_7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ
SEQUENCE 1 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.8%; Score 2032.5; DB 5; Length 1433;
Best Local Similarity 32.6%; Pred. No.2.8e-144;
Matches 471; Conservative 196; Mismatches 395; Indels 385; Gaps 36;

26 CTGTDMKRLRLPASPETHLDMLRHLVYCGQVQVNLLETVLPNTASLSFLQDIOEVQVYL 85
1 CIGTNGMSVAPAREVHYKHLRDRYNTCTYVDSGLLETWQNTIDNLFQIHIREVTGYL 60
86 IAHNQVQVPLQRLIRYGTOLF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
61 ISLYDLQVLLPRLQILRGRTTFKLNKEEAYGLFV-----SFSHMTL 104
141 QLSLSTELKGVLIQNPOLCYODTILMDQYIKASKF-IGITELNRSACHPCSPMC 199
105 ELPLRLDILGSGVFENNYNLCHMKSLNMBEILAPQTSMQYTENFSSPERVCPCHPSC 104
200 KGSRCWGESSEDCSLTRTVACAGCA--RCKGPLPTDCCHEQCAAGCTGPKHSQCLCLH 257
165 EVG-CWEGAHNQRFSKLNCSPOCSGRCFGRPRECCHLFCAGCCTGPTOSDCLCKN 223
258 FNHSGICEHLCPALVTYNTDTFESMPNREGRYTFGASCYTAACPYNVLSTDVGCTLYCPL 317
224 FYDDGVCKQRCPPMQIINPTNYFMEBPDKYAVGATCVAKCP-EHLIKDNGACVRRCPK 282
318 HNGEVLAEDEGTQRCCKSKRCARVCYGLGEMHAREVAVTSANIQEAGCKKIFGSLAF 377
283 GKRPQNSE-----CVPKGVCPKTCPEGIVH-----SDIGNKDKCTIILEGSLIL 329
378 PESFDGDPASNT-----APLPQEOQLVFTLEETIGLYLISAWPDSLPDLSVFONLQ 429
330 DQSFDFGQOYVTFPSGPRYIKIDPRLVLFVSIVKELTGFINIQAHHPNFTTLNYPNLE 389
430 VIRGILHNGAY-SLTLOGISWLGRLSRLREGSLALIHNTHLCFVATVPMDOLFNR 488
390 VVGGRQLKENTLFASVYIVKTSLSKLEKSLKRVNSGIVILENSDLCFVEDIDMSEIKKS 449
489 PHQALLHTANREDECVEGEGLAGHOLCARGHMGPGPTQCVACAHYKDPFECVA 548
450 SHDEVVQKRNATTECHBERGBCSEQSKAGCWGKGPQDCLBCKNVYKKGKCLDSCK--- 506

```

OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

```

OY 549 GLPREY-VNARHCLPCHPECOPONGSVTCFPEADQCAVACAHYKDPFCVACRP----- 601
DB 507 SLRPLYSVDSKTCGDCHQECCKD-----FCYGPENEDNCSSCMVAVDGRCAVACEPTTKAM 561
OY 602 -----SGVKPDLSTYMP1WKFPD----- 618
DB 562 NGTCINCHKTCVGRGPRDITAPDGCISCDKAIIGSDAKIERCLMKDESCPDGYSDYVL 621
OY 619 -EEG----- 621
DB 622 QEEGPLKQLSGKAVCRKCHPRCKCTGYGFHEQFCQECTGYKKGBQCEDECPDYANEE 681
OY 622 --ACOPCPINCT-----HSCVDL-----D-----KCGPAEQ----- 646
DB 682 TRILPCHQEGRCHGLADHDHNECNLTLFEDPDYDNTTTCVANCASHRYKRFPOEA 741
OY 647 -----RASPLTSIVSAVGLLVVLGVFGI---LIRKROQKIRKXTM 687
DB 742 GKIGPYCSADSMQSGRIEPOTVKIVMGSVWALLILLCVFGIAFLVFSRHKKDKDAVGM 801
OY 688 RRLQETELVPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIDGENVK 747
DB 802 TMLAGCEDSEPLRASNVPNLTKRIIKEAIRRGVAGMGAFGRVFKGVMMPEGESVK 861
OY 748 IPVAIKVLRENTSPKANKELIDEAYVMAVGSPIVSRLLGICLTSTVOLVTOLMPYGCIL 807
DB 862 IPVAIKVLMEWSGSSSKFLEBAVIMASVEHPNLKLLAVCMISOMMLIQLMLGCLL 921
OY 808 DHVENRGRIGSODLLNWCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKTIDFGIA 867
DB 922 DYVNNKDKIGSKALLNWSIOIARQMAYLEERLVRDLAARNVLVQPSCVKIVFGIA 981
OY 868 RLILIDETEVADGKVKIKKMALESLIRRFTHQSDVMSGVNLTWELMTREAKRYDILP 927
DB 982 KLILDSDEYPAAGSKMPLKMLALECIRHVFSTSKSDVMAFGITTWELTTGARLEYENVP 1041
OY 928 AREIPDLLEKGERLPEOPICITIDVYIMVCMIDSECRPRERELVSFSEMAADPQRFV 987
DB 1042 AKDVPELEIGHKLPQPDICSLDYVCIILSCVULADADARPTFKQALFAEKARDPGYVL 1101
OY 988 VIONEDLGPASPLDSTFYRSLLEDDMDGLV----- 1018
DB 1102 MI-----PGDKFMRPLSTYNDQDEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152
OY 1019 DAEEYLVPOQGFPCDPAPGAGWVHNHRSSSTRSGGDLTLGLEPSEBEAPRS----- 1073
DB 1153 ETDEYLOPKTRPSIMLPQPSA-----VEPS-DEMPKSLRYCK 1188
OY 1074 -PLAP---SEGAGSDVFDGDLGMAAGKLGSLPTHDPSPLOQRYSEDPVPLPSETDGYVA 1129
DB 1189 DPLKRDDETDGKKEV-----GVGGR-----LNLPLDEDDVLM 1222
OY 1130 PLTSPQPEYVNPQVPRQPSPRGRLPAPARPAGATLERAKTSLPGKNGVVKVDFAFGC 1189
DB 1223 P-TCOSQ---NQS-----TPG---YMDLIGVPA 1243
OY 1190 AVENPEYL-----TPGGAAPQPPPPAPSPAFNLVYWDQDPPERGAPESTFKGT 1240
DB 1244 SVDNPEYLMGSTQAIAGLAQSGM--PHTPP-----PPTTPNGM 1280
OY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

RESULT 8

```

OY 09UK79 PRELIMINARY: PRT; 419 AA.
AC 09UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DB Herstatin.

```

```

GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF17761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FFC1BE347E2D030C CRC64.

```

```

Query Match 26.5%; Score 1802.5; DB 4; Length 419;
Best Local Similarity 86.3%; Pred. No. 1.1e-127;
Matches 340; Conservative 9; Mismatches 38; Indels 7; Gaps 2;

```

```

OY 1 MELALCRWGLLALLPPGAASTOVCTGTDKRLPASPEHLDMLRHLYOGCQVVGSL 60
DB 1 MELALCRWGLLALLPPGAASTOVCTGTDKRLPASPEHLDMLRHLYOGCQVVGSL 60
OY 61 ELTILPTASISPLDIOEVGVYLIANNQVQVFLQRLIVRGQLFEDNYALAVLNG 120
DB 61 ELTILPTASISPLDIOEVGVYLIANNQVQVFLQRLIVRGQLFEDNYALAVLNG 120
OY 121 DPLNNTPTVGTASPGARELQRLSTELIKSGVLIQRNPOLCYODTILMKDOYIANSKF 180
DB 121 DPLNNTPTVGTASPGARELQRLSTELIKSGVLIQRNPOLCYODTILMKDOYIANSKF 180
OY 181 IGITELNRSRACHPSPCKSRCKSGESSEDCOSTLRTVCAGGACRCKGRLPTDCHEQC 240
DB 181 LTLIDTNSRACHPSPCKSRCKSGESSEDCOSTLRTVCAGGACRCKGRLPTDCHEQC 240
OY 241 AAGCTGPRHSCTCLACLNHNSGICELRCPALVTYNTDTFESMPNPEGRTTGASCVTACP 300
DB 241 AAGCTGPRHSCTCLACLNHNSGICELRCPALVTYNTDTFESMPNPEGRTTGASCVTACP 300
OY 301 YNYLSTDVGSCTLVCPPLNNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPPLNNOEVTABDGTORCEKSKPCAR-----GTHSLPPRAAVPVP 355
OY 361 IQEPAGCKKIFGSLAFPESEFGDPASNTAPLP 394
DB 356 LRMQPG--PAHPVLSFLRPSWDLVSAFYSLPAP 387

```

RESULT 9

```

OY 08R2X1 PRELIMINARY: PRT; 367 AA.
AC 08R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DB Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
DB Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.

```

DR EMBL/ BC027080.1; --
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5,5e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

```
QY 889 MALESILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICT 948
DB 1 MALESILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICT 60
QY 949 IDVMIMVKKMMIDSECRPRFRELVSFRRMARDQRFVVIQNEDELGSSPDSFTFYSL 1008
DB 61 IDVMIMVKKMMIDSECRPRFRELVSFRRMARDQRFVVIQNEDELGSSPDSFTFYSL 120
QY 1009 LEDDDMGDLVDAEELVPOQGFPCDPAPAGAWHHHRSSSTRSGGDLTLGLEPSEE 1068
DB 121 LEDDDMGDLVDAEELVPOQGFPCDPAPAGAWHHHRSSSTRSGGDLTLGLEPSEE 180
QY 1069 EAPRSPLAPSGAGSDVFDGDLGCAKAGLQSLPTHDPSPLOVSEDPVLPSETDGYV 1128
DB 181 EAPRSPLAPSGAGSDVFDGDLGCAKAGLQSLPTHDPSPLOVSEDPVLPSETDGYV 240
QY 1129 APLTCSPOPEYVNDPVRPQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVFAFG 1188
DB 241 APLTCSPOPEYVNDPVRPQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVFAFG 300
QY 1189 GAVENPEYLTPQCGAAPQHPHPPAPSPAFDNLVYWDQDPREGAPSPSTFKGPTAENEY 1248
DB 301 GAVENPEYLTPQCGAAPQHPHPPAPSPAFDNLVYWDQDPREGAPSPSTFKGPTAENEY 360
QY 1249 LGLDVPPV 1255
DB 361 LGLDVPPV 367
```

```
RESULT 10
086712 PRELIMINARY; PRT; 729 AA.
ID 086712
AC 086712;
DT 01-NOV-1996 (Tremblrel. 01, Created)
RT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Polypeptide.
OS Polypeptide.
GN Polypeptide.
OC Avian rous-associated virus type 1.
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxId=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; --
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84DP6914EFE1D63 CRC64;
```

Query Match 25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4,2e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

```
QY 569 PONGSVTCFGEADQVCAHYKDPFCVACRPSGVKDLSTMPYWKPFDEGACQPCPI 628
DB 141 PESTATPPTGP--DHCKMCAHFTIDGPHCVKACPAVLGENDTL-VMKADANAVQQLCHP 197
QY 629 NCTHSCVLDLADKGCAPAEORASPLTIVSAVV-GILLVYVAVGFGILLKROOKIRKYM 687
DB 198 NCTHSCVLDLADKGCAPAEORASPLTIVSAVV-GILLVYVAVGFGILLKROOKIRKYM 253
QY 688 RLLLOETELVPLTPSGAMPNOQMRILKETELRKVYVLSGSAFGTVYKGIWIPGGENYK 747
DB 254 RLLLOETELVPLTPSGAMPNOQMRILKETELRKVYVLSGSAFGTVYKGIWIPGGENYK 313
QY 748 IPVAIKVLENTSPANKELIDEAIVMAVGVSPYSLGLCTISVTLVTLQMPYGCIL 807
DB 314 IPVAIKVLENTSPANKELIDEAIVMAVGVSPYSLGLCTISVTLVTLQMPYGCIL 373
QY 808 DHVENRGRGLGSODLLNMCQIAGKMSYLEVRLVHRDLAARNVLYKSPNHVKITDFGLA 867
DB 374 DHVENRGRGLGSODLLNMCQIAGKMSYLEVRLVHRDLAARNVLYKSPNHVKITDFGLA 433
QY 868 RLDDIDETEHADGAKVPIKMALESILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGI 927
DB 434 RLDDIDETEHADGAKVPIKMALESILRRRTTHQSDVMSYGVTTWELMTFGAKPYDGI 493
QY 928 AREIPDLLEKGERLPPPICTIDVYMTIMVKKMIDSECRPRFRELVSFRRMARDQRFV 987
DB 494 AREIPDLLEKGERLPPPICTIDVYMTIMVKKMIDSECRPRFRELVSFRRMARDQRFV 553
QY 988 VIQ-NEIDGPASPLDSTFYSRLLEDGDLVDAEELVPOQGFPCDPAPAGAGMHNH 1046
DB 554 VIQ-NEIDGPASPLDSTFYSRLLEDGDLVDAEELVPOQGFPCDPAPAGAGMHNH 598
QY 1047 HRSSSTRSGGDLTLGLEPSEEAAPRSP-----APSEGASDVFDGDLGCAKAGLQSL 1101
DB 599 HRSSSTRSGGDLTLGLEPSEEAAPRSP-----APSEGASDVFDGDLGCAKAGLQSL 631
QY 1102 PTHPSPLOVSEDPVLPSET--DGVAAPLTCSPOPEYVNDPVRPQPPSPREGPLPA 1159
DB 632 PTHPSPLOVSEDPVLPSET--DGVAAPLTCSPOPEYVNDPVRPQPPSPREGPLPA 675
QY 1160 ARPAGATLERAKTILSPKNGVYKDV-----AFGAVENPEYL 1197
DB 676 ARPAGATLERAKTILSPKNGVYKDV-----AFGAVENPEYL 715
```

```
RESULT 11
086714 PRELIMINARY; PRT; 567 AA.
ID 086714
AC 086714;
DT 01-NOV-1996 (Tremblrel. 01, Created)
RT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-erbB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxId=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; --
DR HSSP; P11362; 1F6K.
DR InterPro; IPR000719; Euk_pkinase.
```

DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; kinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR Tyrosine-protein kinase.
 FT NON TER
 SO SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 4e-121;

Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCAAHYKDPFCVACPSGVKBDLSYMIWFPDEGACOPPCINCHSCVDL 637
 DB 1 GP--DHCKCAHFIIDGPHCVKACPGVIGENDTL-VKXADANAACOLCHPNCIGCKGP 57
 QY 638 DDKGCPAEQASPLTSIVSAVY-GILLVVLGVGVGILIKRQOKIRKTYMRLLQETEL 696
 DB 58 GLEGCP---NSKTSIAAGVVGILLVVGILGILYRR-HYRKRTLRLLQEREL 113
 QY 697 VEPLTPSGAMPNQAQRIKTELKRVKVGSGAFVYKGIWPDGENVKIPVAIKYLR 756
 DB 114 VEPLTPSGEAPNQAHLRIKETEFKVKVLSGAFGVYKGLMIDEGEKVIPVAIKELR 173
 QY 757 ENTSPKAKETLDEYVYVAGVSPVYSRLGICLTSTVOLVQLMPYCGILDHVENKGR 816
 DB 174 EATSPKAKETLDEYVYVAGVSPVYSRLGICLTSTVOLVQLMPYCGILDHVENKGR 233
 QY 817 LGSODLNMCMQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKTITDGLALDIDETE 876
 DB 234 IGSQVLNMCQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKTITDGLALDIDETE 293
 QY 877 YHADGKPIKMALESLRRFTYQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
 DB 294 YHAEKGKPIKMALESLRRFTYQSDVSYGVTVWELMTFGSKPYDGIIPAREIPDLLE 353
 QY 937 KGERLPQPICTIDVYMIWVCMMDSECRPRELVEFSRMAADPQRFVITO-NEGL 995
 DB 354 KGERLPQPICTIDVYMIWVCMMDSECRPRELVEFSRMAADPQRFVITO-NEGL 413
 QY 996 PASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFCCDDPAPAGAMVHRRSSSTSG 1055
 DB 414 LPSPLDSTFYRSLLEDMDGLVDAEYLVPOQGFCCDDPAPAGAMVHRRSSSTSG 454
 QY 1056 GGDLLTGLPSEBEAPRSPPL-----APSEGAGSDVFDGLGKAAGLQSLFTHDPSPLQ 1110
 DB 455 -----SRTPLLSLSLATSNSATNCID-----RNGCGHVRDEDSFVQ 491
 QY 1111 RYSEDPVPLPSET--DGTVAPLITCSPOPEYVNOQDVAPQPSPREGLPAPARAGATLE 1168
 DB 492 RYSSDPVPLPSET--DGTVAPLITCSPOPEYVNOQDVAPQPSPREGLPAPARAGATLE 526
 QY 1169 RAKTLSPGKNGVYKDVF-----AFGAVENPEYL 1197
 DB 527 -----TAMVQNGIYNNISLTALSKLPMDSRYQNSHSTADNPEYL 566

RESULT 12

Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0; 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 GN Hypochemical 44.7 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBAJ databases.
 DR EMBL; AF318349; AAL55856.1; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR00719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Hypochemical protein.
 SO SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 8.9e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRFTYQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 948
 DB 1 MALESILRRFTYQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60
 QY 949 IDVYMIWVCMMDSECRPRELVEFSRMAADPQRFVITONEDLGPAASPLDSTFYRSL 1008
 DB 61 IDVYMIWVCMMDSECRPRELVEFSRMAADPQRFVITONEDLGPAASPLDSTFYRSL 120
 QY 1009 LEDDDMGDLVDAEYLVPOQGFCCDDPAPAGAMVHRRSSSTSGGDLTLGLEPSEE 1068
 DB 121 LEDDDMGDLVDAEYLVPOQGFCCDDPAPAGAMVHRRSSSTSGGDLTLGLEPSEE 180
 QY 1069 EAPRSLPSPGAGSDVDGDLGMAAGLQSLPTHDSPLQRYSEDPVPLPSETDGV 1128
 DB 181 EAPRSLPSPGAGSDVDGDLGMAAGLQSLPTHDSPLQRYSEDPVPLPSETDGV 240
 QY 1129 APLTCSPOPEYVNOQDVAPQPSPREGLPAPARAGATLERAKTSPGKNGVYKDVAFG 1188
 DB 241 APLTCSPOPEYVNOQDVAPQPSPREGLPAPARAGATLERAKTSPGKNGVYKDVAFG 300
 QY 1189 GAVENPEYLTPOGGAAP-----HPPA---FSPAFDNL 1220
 DB 301 GAVENPEYLTPOGGAALPTLLPSAOPSTTITGTRTQSGCLHAPSKGHLRQGTOST 360
 QY 1221 YYMD-QDPPER-----GAPSTFKGTPTAEN 1245
 DB 361 WMTQCCEPQVRRSPVSSGREGLTSGAKIRMEGPTTSGTCHARN 410

RESULT 13

Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895; 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag-v-erb-A-v-erb-B protein.
 GN Gag-v-erb-A-v-erb-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OC NCBI_TaxID=11861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90206603; PubMed=1969616;
 RA Brunkin A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.,
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B.";

```
RL Oncogene 5.15-24(1990)
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; X52209; CA36459.1; -
DR EMBL; X52211; CA36459.1; JOINED.
DR HSSP; P10828; NUL.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR000536; Hormone_rec_1ig.
DR InterPro; IPR001723; Slc4hrm_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_Csteroi.
DR Pfam; PF00104; hormone_rec.1.
DR Pfam; PF00069; kinase.1.
DR Pfam; PF00105; zf_C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR Prodom; PD000001; Euk_kinase.1.
DR Prodom; PD000035; Znf_Csteroi.1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TykC; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_KINASE_ATP; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS01009; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6.6e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEEGRVQLPRE-YVNAR-HCLP-----CHPEQ 568
DB 354 IEKQSEYLAFFHYIRGNITPHFWSKLMKADLMICAYASRFLHMKVEPEL 413
QY 569 PONGSVTCFPEADQVACAHYKDPFCVACPSGVKPDLSYMDIWKFPDEGACQPCPI 628
DB 414 PQE-----VGP--DHCKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCQLCHP 465
QY 629 NCHSCVDLDDKCGPAQRASPLTISAVV-GILLVVVLGVVGLIKRQOKIRKXTM 687
DB 466 NCRGCGPGLGCP--NGSKTPSIAAGVGLICLVVGLIGLYLRR-HIVRKTL 521
QY 688 RRLLOETELVEPLTPSGAMPNOAOMRIKETELRKVVLGSGAFGYKGIWDGENVK 747
DB 522 RRLLOEELVEPLTPSGEARQAHRLIKETEFKVKVGLGAFGYKGLMPEGEKVT 581
QY 748 IPVAIKVLRNTSEPKANKELIDEAYVAVGSPYVSRLLGICLTSTVOLVQLMPYGCCL 807
DB 582 IPVAIKELREATSPKANKELIDEAYVAVSDNPHVCLLGLICLTSTVOLITQLMPYGCCL 641
QY 808 DHVRENGRGLGSDLLWMCQIAGMSYLEVRLVHRDLAARNVLYKSPNHVKTITFGLA 867
DB 642 DYIREHNDNGSOYLWMCQIAGMSYLEERHMHVRLAARNVLYTPQHVKITDFGLA 701
QY 868 RLIDIDETEHADGKVPKIMMALESILRRRFTHQSDVMSYGVTVLMETFGAPYDGP 927
DB 702 KQCGADEKEVHAEGKVPKIMMALESILHRIYTHQSDVMSYGVTVLMETFGSPYDGP 761
QY 928 AREIPDLLEKGERLPQPPCTIDVYIMVKKCMWIDSECRPRFELVSEFSRMADPQRFV 987
DB 762 ASBISSTLEKGERLPQPPCTIDVYIMVKKCMWIDSECRPRFELVSEFSRMADPQRFV 821
QY 988 VIO-NEDELGASPLDSTFYVSLLEDDMDGLVDAEETLVPOQGFCCDPAPAGAGMYHR 1046
DB 822 VIOGDEKMLPSPTDSKFYRFLMEEDMDIIVDAEYLVPHQGF----- 866
QY 1047 HRSSSTRSGGDLTLGLPESEEARSPPLAPSEGAGSDVFDGDLGMAAGKGLSLPTHP 1106
DB 867 -NSPST-----SRTPLLSSLSATSN-----NSATYCIDRNGH-- 898
```

```
QY 1107 SPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYVQNPVRPQPSREGPLPAPAGAT 1166
DB 899 -----PVREDGFL-----PAPEYVQ--LMPKPESTAVQNIYVISLT 936
QY 1167 -LERAKTLSPQNGVKKVFAFGAVENPEYL 1197
DB 937 AISKLPMSRYQN-----SHSTAVDNPETL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMBLrel. 01. Created)
DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20. Last annotation update)
DE Avian erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
Virus; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase.1.
DR Prodom; PD000001; Euk_kinase.1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS01009; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CA078AF4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.2e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQVACAHYKDPFCVACPSGVKPDLSYMDIWKFPDEGACQPCPINCNSVDL 637
DB 1 GP--DHCKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCQLCHPCTGCKGP 57
QY 638 DDKCPAQRASPLTISAVV-GILLVVVLGVVGLIKRQOKIRKXTMRRLLOETEL 696
DB 58 GLEGCP--NGSKTPSIAAGVGLICLVVGLIGLYLRR-HIVRKTLRRLLOEREL 113
QY 697 VEPULTPSGAMPNOAOMRIKETELRKVVLGSGAFGYKGIWDGENVKIPVAIKVLR 756
DB 114 VEPULTPSGEAPNOAHRLIKETEFKVKVGLGAFGYKGLMPEGEKVTIPVAIKELR 173
QY 757 ENTSPKANKELIDEAYVAVGSPYVSRLLGICLTSTVOLVQLMPYGCCLDHVRENGR 816
DB 174 EATSPKANKELIDEAYVAVSDNPHVCLLGLICLTSTVOLITQLMPYGCCLDYIREHND 233
QY 817 LGSODLLWMCQIAGMSYLEVRLVHRDLAARNVLYKSPNHVKTITFGLARLIDIDETE 876
DB 234 IGSQYLLWMCQIAGMSYLEERLIVHRDLAARNVLYKSPDHVKTITFGLKQGLADKE 293
QY 877 YHADGKVPKIMMALESILRRRFTHQSDVMSYGVTVLMETFGAPYDGPAREIPDLLE 936
DB 294 YHAEKGVKVPKIMMALESILHRIYTHQSDVMSYGVTVLMETFGSPDGPASSISSTLE 353
QY 937 KGERLPQPPCTIDVYIMVKKCMWIDSECRPRFELVSEFSRMADPQRFVIO-NEDELG 995
DB 354 KGERLPQPPCTIDVYIMVKKCMWIDSECRPRFELVSEFSRMADPQRFVIOGDERMH 413
```



```

QY 996 PASPLDSTFYRSLLIEDDDMDGLVDAEEYLPQGFCCPDPAAGAGMHHRRSSSTRSG 1055
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 414 LPSPTDSEFYRTLMEEEMEDIVDADEYLVPHQGF-----NGST----- 454
QY 1056 GGDLLTLGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGKAAGLQSLPTHDPSPLQ 1110
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 455 -----SRTPLLSLSATSNNSATNCIDRNGC-----H----- 481
QY 1111 RYSDPTVPLPSETDGYVAPLTCSPQPEYVNPDPVRPPSPREGPLPAARAGAT-LER 1169
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 482 -----PVREDEGLP-----PAPRYVNO-LMPKKPSTAMVOIQINYISLTATSK 523
QY 1170 AKTSPGKGVKDVFAFGAVENPEYL 1197
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 524 LPMDSRYGN-----SHSTAVDNEPYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT: 655 AA.
AC Q9WVF5.
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Mahtle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Mahtle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Konno H., Aachi J., Fukuda S., Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana I., Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flaischmann W., Gaasterland T., Gissi C., Ring B., Kochia H., Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J., Schmitt L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G., Blake J., Boffelli P., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontecki S., Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DB EMBL; AF124513; AAD44149.1; -.
DB EMBL; AF275366; AAG28047.1; -.
DB EMBL; AF275364; AAG28047.1; JOINED.
DB EMBL; AF275365; AAG28047.1; JOINED.
DB EMBL; AK004944; BAB23688.1; -.
DB EMBL; AK004883; BAB23641.1; -.
DB EMBL; AK004911; BAB23662.1; -.
DB MGI; MGI:95294; Egfr.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.4%; Score 1524.5; DB 11; Length 655;
Best Local Similarity 44.1%; Pred. No. 2,1e-106;
Matches 280; Conservative 100; Mismatches 240; Indels 15; Gaps 5;

QY 11 LLLALLPPGAA--STOVCTGDMKLRLPASPEYTHLMDRLHYOGCOVGNLELYTPN 68
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 14 LLLTALCAAGALAEKXKVCOGTSNRLTQGTGFEDHFLSLQBMVYNNCEVGLNLEITYVQRN 73
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 69 ASLSFLDIOIEYQGVYLIANQVROVPLQRLRIYVGTQLFEDNYALAVDNDPLNNTTP 128
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 74 YLSFLKTIQIEVAGVLIATLVTERIPLNQIIRGNALYENTVALATLSN----- 124
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 VTGASPGRLRELQSLREILKGVLIQRNPOLCYOPTLWMDQYIKANSKFIGITELNR 168
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 125 -YGNKRTGLREIPMLNLOEILGAVFSNNPLCNMDITIQMDIYQVNFMSNSMDLQSH 183
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 189 SRACPCSPMKGSKRCWSESSBDCSLTRTYCAGCA-RCKGPLPTDCHECAAGCTGP 247
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 184 PSSCPKCPSCPCNGSCWCGEENCKLTKITCAQGCSSHRCSRSPSDCHNCAGACTGP 243
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 248 KHSDDLALHFRHSGICGLHCPALVTYNTDFESHPNBEGRITFGASCTYACPYNYLSD 307
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 244 RESDCLVQKQFQDEATKDTCPPLMLYNPTTYQMDVNPBGKYSFGATCVKCKPRNVYVD 303
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 308 VGSCTLVPLPHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATVSANIOEFAC 367
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 304 HGSCYRACGPDYEV-EDDGIRKCKCKGCPCKVCGNGIGGFKOTLSINAINIKFKYC 362
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 368 KKIIFGSLAFLPESFDDPSANTAPLQPEQLQVFETLEBITGYLYTISAMPDSLPLDSVFQ 427
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 363 TALSDDLHLPLVAFKDGSTFRTPPLDPRELEIKTVKEITGFLLIQAMPDNTDLHAFEN 422
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 428 LOVIRGRILHNGAVSLTQIGISWLGRLSRELSCGLALIHNNTHLCFVHTVPNDQLFR 487
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 423 LEIRIGRTRKHQGFSLAVVGLNITSLSGLRSLSKEISDGVIIISGNRNLCYANTINNKKFLG 482
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 488 NPHQALHTARPREBECEVGEGLACHQLCARGHGCMWGPPOCNCSQPLRGQCEVEECRYL 547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 483 TPNQTKTKMNAEADKCAVNVHVCNPLCSSBCCWCPPEPDVSCCNVNSRGRECEVCNLT 542
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 548 QGLPREYVNAHCLCPHPCPCPQNGSVTCFGEADQCYACAHYKDPPEFCVAPCPSPGVKD 607
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 543 EGEPEPEFVENSECICHCHECLPQANINICTGNGPNCIQCAHYIDGPHCVKCPAGIMGE 602
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 608 LSYMPIKFPDEBGCQCPCPINCHSCVDLDDKGC 642
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 603 NNTL-VWKYADANNVCHLCHANCTYGACGPGLOGC 636
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: July 22, 2003, 09:00:28
 Job time : 53.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)

4403.399 Million cell updates/sec

Title: SEQ4-149-163-12

Perfect score: 6810
Sequence: 1 MELALCWMGLLLALPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6740	99.0	1255	21	AAV92620
2	6740	99.0	1255	22	AAE12130
3	6740	99.0	1255	22	AAU60167
4	6740	99.0	1255	23	AAU74545
5	6734	98.9	1255	17	AAW01111
6	6734	98.9	1255	20	AAW92406
7	6734	98.9	1255	21	AAAB21198
8	6734	98.9	1255	21	AAAB4780
9	6734	98.9	1255	22	AAAB5458
10	6734	98.9	1255	22	AAAB8267

11	6734	98.9	1255	23	AAE24067
12	6734	98.9	1255	23	AAE20479
13	6734	98.9	1255	23	AAW51143
14	6734	98.9	1255	23	AAU77114
15	6691	98.3	1433	14	AAAB39568
16	6570	96.5	1223	23	AAU98923
17	6417	94.2	1200	21	AAAB21208
18	5945.5	87.3	1256	23	AAAB21199
19	5945.5	87.3	1256	23	AAW51144
20	5925.5	87.0	1256	23	AAAB21206
21	5925.5	87.0	1256	22	AAAG2860
22	5925.5	87.0	1256	23	AAW51151
23	4820	70.8	919	23	AAAB21203
24	4820	70.8	919	23	AAW51148
25	4075.5	59.8	920	23	AAW51152
26	4075.5	59.8	926	23	AAW51153
27	3704	54.4	712	21	AAAB21204
28	3704	54.4	712	23	AAW51149
29	3558	52.2	782	18	AAW19764
30	3556	52.2	653	21	AAAB21200
31	3556	52.2	653	23	AAW51145
32	3518	51.7	645	22	AAAB50408
33	3518	51.7	645	22	AAAB51593
34	3453	50.7	951	21	AAAB44993
35	3350	49.2	624	11	AAAB08222
36	3160	46.4	1210	21	AAAB19259
37	3160	46.4	1210	23	AAV50616
38	3160	46.4	1210	23	AAE23019
39	3160	46.4	1210	23	AAW50768
40	3158	46.4	1210	22	AAAB6420
41	3119	45.8	1210	22	ABP51768
42	3084	45.3	583	23	AAE20483
43	3084	45.3	587	23	AAE20481
44	3083	45.3	589	23	AAE20484
45	3083	45.3	600	23	AAE20482

ALIGNMENTS

RESULT 1	AAV92620	standard; Protein; 1255 AA.
ID	AAV92620	
XX	AAV92620;	
AC	10-AUG-2000	(first entry)
XX		
XX		
DE	Human heregulin 2 (Her2).	
XX		
KW	Her2; vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
OS	Homo sapiens.	
XX		
XX		
PH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= "N-terminal"
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= "insertion region"
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 1011..1235
 FT /label= C-terminal_domain
 FT
 PN WO200020027-A2.
 XX 13-APR-2000.
 PF 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 PA Steinaa L, Mouritsen S, Nielsen KG, Haaning J, leach D, Dalum I,
 PI Gautam A, Birk P, Karlsson G;
 XX WPI: 2000-349917/30.
 DR N-PSDB; AAA09455.
 DR
 XX Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 PS Claim 62; Page 193-198; 220pp; English.
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/preast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 CC XX

SQ Sequence 1255 AA:

Query Match 99.0%; Score 6740; DB 21; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALLPPGAATQVCTGDMKRLPASFETHLDMRLHYOGGVQGNL 60
 DB 1 MELALCRWGLLALLPPGAATQVCTGDMKRLPASFETHLDMRLHYOGGVQGNL 60
 QY ELTVLPPTNASLFLQDIQEVGVYLIHNOVQVPLQRIYRGTOLEFDNYALAVLDNG 120
 DB ELTVLPPTNASLFLQDIQEVGVYLIHNOVQVPLQRIYRGTOLEFDNYALAVLDNG 120
 QY 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
 DB 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
 QY 122 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
 DB 122 DPLNNTPTVPGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
 QY 181 IGTIELNRRACHPCSPMKCKGRCESEDCOSLTRTYCAGGACARCKPLPTDCHEQC 240
 DB 181 IGTIELNRRACHPCSPMKCKGRCESEDCOSLTRTYCAGGACARCKPLPTDCHEQC 240
 QY 181 LRLITNRRACHPCSPMKCKGRCESEDCOSLTRTYCAGGACARCKPLPTDCHEQC 240
 DB 181 LRLITNRRACHPCSPMKCKGRCESEDCOSLTRTYCAGGACARCKPLPTDCHEQC 240
 QY 241 AAGCTGPKASDCLACIHFHSGICELHCPALVTYNTDESPNPNRGRTPFGASCYTACP 300
 DB 241 AAGCTGPKASDCLACIHFHSGICELHCPALVTYNTDESPNPNRGRTPFGASCYTACP 300
 QY 241 AAGCTGPKASDCLACIHFHSGICELHCPALVTYNTDESPNPNRGRTPFGASCYTACP 300
 DB 241 AAGCTGPKASDCLACIHFHSGICELHCPALVTYNTDESPNPNRGRTPFGASCYTACP 300
 QY 301 YNYLSTDVSGCTLVCEPLHNOEVTABDGTORCEKSPCARVCYGLGMEHLREVAATYSAN 360
 DB 301 YNYLSTDVSGCTLVCEPLHNOEVTABDGTORCEKSPCARVCYGLGMEHLREVAATYSAN 360
 QY 301 YNYLSTDVSGCTLVCEPLHNOEVTABDGTORCEKSPCARVCYGLGMEHLREVAATYSAN 360
 DB 301 YNYLSTDVSGCTLVCEPLHNOEVTABDGTORCEKSPCARVCYGLGMEHLREVAATYSAN 360
 QY 361 IOEPFACCKKI FGSFLAFLPESFDGDPASNTAPLQPELOVETLEETGYLYISAWPDSL 420
 DB 361 IOEPFACCKKI FGSFLAFLPESFDGDPASNTAPLQPELOVETLEETGYLYISAWPDSL 420
 QY 361 IOEPFACCKKI FGSFLAFLPESFDGDPASNTAPLQPELOVETLEETGYLYISAWPDSL 420
 DB 361 IOEPFACCKKI FGSFLAFLPESFDGDPASNTAPLQPELOVETLEETGYLYISAWPDSL 420
 QY 421 DLSVFONLQVIRGRILHNGAYSILTLQGLISWLGRLSLELSGLALIHNNHLCFVHTV 480
 DB 421 DLSVFONLQVIRGRILHNGAYSILTLQGLISWLGRLSLELSGLALIHNNHLCFVHTV 480
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHGKMGPTQVNCQSFRLGQEC 540
 DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHGKMGPTQVNCQSFRLGQEC 540
 QY 541 VEECKVLOGLPREYVNAHRCLEPCHPECQONSSVTCFGBADQCAVCAHYKDPFCVAPC 600
 DB 541 VEECKVLOGLPREYVNAHRCLEPCHPECQONSSVTCFGBADQCAVCAHYKDPFCVAPC 600
 QY 541 VEECKVLOGLPREYVNAHRCLEPCHPECQONSSVTCFGBADQCAVCAHYKDPFCVAPC 600
 DB 541 VEECKVLOGLPREYVNAHRCLEPCHPECQONSSVTCFGBADQCAVCAHYKDPFCVAPC 600
 QY 601 PSGVPPDLSYMPIMWFPDEEGACOPCPINCHTSVDLDDKGPABGRASPILTSYSAVVG 660
 DB 601 PSGVPPDLSYMPIMWFPDEEGACOPCPINCHTSVDLDDKGPABGRASPILTSYSAVVG 660
 QY 661 ILVVVVLGVVFGILIKRRQOKIRKTYMRRLQETELVBEPLTPSGAMPNQAKRLKETEL 720
 DB 661 ILVVVVLGVVFGILIKRRQOKIRKTYMRRLQETELVBEPLTPSGAMPNQAKRLKETEL 720
 QY 721 RRVKVLGSGAFETVYKGIWIPDGENVKIPVAIKVIRENTSPANKKEILDEAVVMAGVGP 780
 DB 721 RRVKVLGSGAFETVYKGIWIPDGENVKIPVAIKVIRENTSPANKKEILDEAVVMAGVGP 780

```
QY 781 YSRLLGICLTSTVOLVLTQMPYGCCLLDHRENRGRIGSODLNMWQIAKMSYLEDDR 840
DB 781 YSRLLGICLTSTVOLVLTQMPYGCCLLDHRENRGRIGSODLNMWQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEXHAOGGKPIKMMALLESILRRFT 900
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEXHAOGGKPIKMMALLESILRRFT 900
QY 901 HQSDVMSYGVTVLWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVACMM 960
DB 901 HQSDVMSYGVTVLWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVACMM 960
QY 961 IDSECRPRRELIVSEFSRMARDQRFVVIQNEIDGSPSLDSTRYRSLLEDDMDGLVDA 1020
DB 961 IDSECRPRRELIVSEFSRMARDQRFVVIQNEIDGSPSLDSTRYRSLLEDDMDGLVDA 1020
QY 1021 EEYLVPOQGFCCPPAPGAGAMVHHRSSSTSGGDDLTLGLEPSESEAPRSLAPSEG 1080
DB 1021 EEYLVPOQGFCCPPAPGAGAMVHHRSSSTSGGDDLTLGLEPSESEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPOPEYV 1140
DB 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPOPEYV 1140
QY 1141 NQDVPVPPQPSPPREGPLPAARPAATLEBAKTLSFGKGVYKDYFAFGAVENPEYLTPO 1200
DB 1141 NQDVPVPPQPSPPREGPLPAARPAATLEBAKTLSFGKGVYKDYFAFGAVENPEYLTPO 1200
QY 1201 GGAAPPPHPPAPFADNLVYWDQDPPERGAPSTKGTPTAENPEYLGIDVPV 1255
DB 1201 GGAAPPPHPPAPFADNLVYWDQDPPERGAPSTKGTPTAENPEYLGIDVPV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
AC AAE12130;
DT 18-DEC-2001 (first entry)
DE Human tyrosine kinase-type receptor, HER-2.
XX
XX Therapeutic compound, major histocompatibility complex; vaccine;
XX antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
XX adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
XX antigen presenting cell; human; tyrosine kinase-type receptor.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Region 774..782
XX FT /note="Antigenic epitope"
XX PN WO200168677-A2.
XX PD 20-SEP-2001.
XX PF 16-MAR-2001; 2001WO-US40328.
XX PR 16-MAR-2000; 2000US-0527487.
XX (GEN2 ) GENZYME CORP.
XX PA
XX PI Nicolette CA;
XX DR MPI: 2001-616284/71.
XX DR N-PSDB; AADI9731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
XX for use in adoptive immunotherapy, has enhanced binding to major
XX histocompatibility molecules and enhanced immunoregulatory properties
XX
```

```
XX
XX Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
XX Sequence 1255 AA;
XX
XX Query Match 99.0%; Score 6740; DB 22; Length 1255;
XX Best Local Similarity 99.0%; Pred. No. 0;
XX Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
QY 1 MEIAALCRWGLLALLPFGAASSTOVCTGTMKRLPASPEHMLRLHYGCGVOVQNL 60
DB 1 MEIAALCRWGLLALLPFGAASSTOVCTGTMKRLPASPEHMLRLHYGCGVOVQNL 60
QY 61 ELTYLPFNASLSFLQDIQEVQGYVLIHANOVRQVPLQRLIRVGTQJFEDNYALAVLDNG 120
DB 61 ELTYLPFNASLSFLQDIQEVQGYVLIHANOVRQVPLQRLIRVGTQJFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGRLREQLRSITELKGVLIQNNPOLCYDTILMWDIFHKNQOLA 180
DB 121 DPLNNTPTVYGASPGGRLREQLRSITELKGVLIQNNPOLCYDTILMWDIFHKNQOLA 180
QY 181 IGITELNRSRACHPCSPMGKSGRCKGSSSEDCOSTRTVCAGGACRCKGPLPTCCHEQC 240
DB 181 LTLIDNTRSRACHPCSPMGKSGRCKGSSSEDCOSTRTVCAGGACRCKGPLPTCCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMPNPEGRTYFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMPNPEGRTYFGASCVTACP 300
QY 301 YNLTSTDVGSCTIVCPAHNDEVTAEDTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNLTSTDVGSCTIVCPAHNDEVTAEDTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IOEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEOLQVEETLEITGYLYISAMPDLP 420
DB 361 IOEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEOLQVEETLEITGYLYISAMPDLP 420
QY 421 DLSVFQNLQVIRRIIHNGAVSLTLOGLGISWLGSLRLSRLSGSLALIHNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRRIIHNGAVSLTLOGLGISWLGSLRLSRLSGSLALIHNTHLCFVHTV 480
QY 481 PMDQLFRNPHQALLHFRANPEDECVEGLACHOLCARHGCMGPGPTCCVNCQFLRQEC 540
DB 481 PMDQLFRNPHQALLHFRANPEDECVEGLACHOLCARHGCMGPGPTCCVNCQFLRQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPEQOPNGSVTCGPAPDQVCAAHKDPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHPEQOPNGSVTCGPAPDQVCAAHKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIMKFPDEBGACQPCPINTHSCVDLDKGCAPAEQASPLTISVAVNG 660
DB 601 PSGVKPDLSTYMPIMKFPDEBGACQPCPINTHSCVDLDKGCAPAEQASPLTISVAVNG 660
QY 661 ILLVTVLVGVFGLIRROOKIRKTYMRRLQSTELVEPLTSGAMPNQAQMRILKXTEL 720
DB 661 ILLVTVLVGVFGLIRROOKIRKTYMRRLQSTELVEPLTSGAMPNQAQMRILKXTEL 720
```

```

Qy 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
    |||||
Db 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
Qy 781 YVSRLLGICLTSTYQVLTQMLPYPGCLLDHRENRGRIGSODLLMWCMQIAKMSYLEDR 840
    |||||
Db 781 YVSRLLGICLTSTYQVLTQMLPYPGCLLDHRENRGRIGSODLLMWCMQIAKMSYLEDR 840
Qy 841 LVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEYHADGKVPIMMALESTLRRFT 900
    |||||
Db 841 LVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEYHADGKVPIMMALESTLRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVXCM 960
    |||||
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVXCM 960
Qy 961 IDSECRFRFRELVESEFRMARDPQRFVYIQMEDLGPPSLDSTFYRSLLEDMDGLVDA 1020
    |||||
Db 961 IDSECRFRFRELVESEFRMARDPQRFVYIQMEDLGPPSLDSTFYRSLLEDMDGLVDA 1020
Qy 1021 EBYLVPOQGFPCPPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEBEARSPPLASEG 1080
    |||||
Db 1021 EBYLVPOQGFPCPPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEBEARSPPLASEG 1080
Qy 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQBEYV 1140
    |||||
Db 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQBEYV 1140
Qy 1141 NOPRVRRQPPSPRGRPLPAARPAATLEBRATLSPGKNGVYKDYFAFGAVENPEYLTPO 1200
    |||||
Db 1141 NOPRVRRQPPSPRGRPLPAARPAATLEBRATLSPGKNGVYKDYFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAPAFDNLVYWDODPPERGAPSTFGKPTJANPEYLGADV 1255
    |||||
Db 1201 GGAAPQHPPPAPAFDNLVYWDODPPERGAPSTFGKPTJANPEYLGADV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein: 1255 AA.
AC
AA B60167;
DT
03-APR-2001 (first entry)
DE
HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
    antibody.
XX
OS Homo sapiens.
XX
XX Synthetic.
XX
PN MO200100244-A2.
XX
04-JAN-2001.
XX
PD
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Ericsson S, Schwall R;
XX
DR WPI, 2001-061962/07.
XX
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
    receptor and does not respond to an anti-ErbB antibody, comprises
    conjugating the antibody to a maytansinoid -
XX

```

```

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
    administering a conjugate of anti-ErbB antibody with a maytansinoid. In
    particular, the antibody is directed against ErbB2 (also known as HER2
    CC and p185neu). The method is particularly useful in the treatment of
    CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
    CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
    CC
SQ Sequence 1255 AA;
    Query Match 99.0%; Score 6740; DB 22; Length 1255;
    Best Local Similarity 99.0%; Pred. No. 0;
    Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
Qy 1 MELAALCRKGLLALLPPGAASVQVCTGTMKRLPASPETHMLRLHYGCGVQGNL 60
    |||||
Db 1 MELAALCRKGLLALLPPGAASVQVCTGTMKRLPASPETHMLRLHYGCGVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHQVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
    |||||
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHQVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVYGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
    |||||
Db 121 DPLNNTTPVYGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILMKDQYIKANSKF 180
Qy 181 IGITELRSRACHSCSPMKCSRWGSSSEBCCSLTTCVCGGCAKCGPLPTCCHEQC 240
    |||||
Db 181 IGITELRSRACHSCSPMKCSRWGSSSEBCCSLTTCVCGGCAKCGPLPTCCHEQC 240
Qy 241 AAGCTGPKHSDDLACLFHNSGICELHCPALVTNTTFESMPNPEGRTYGASCVTACP 300
    |||||
Db 241 AAGCTGPKHSDDLACLFHNSGICELHCPALVTNTTFESMPNPEGRTYGASCVTACP 300
Qy 301 YNYLSTDVGSCTVLCPLHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
    |||||
Db 301 YNYLSTDVGSCTVLCPLHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 IOEPAGCKKIFGSLAFLPESPDGDPASNTAPLOEOLOVFETLEITGYLISAMPDLP 420
    |||||
Db 361 IOEPAGCKKIFGSLAFLPESPDGDPASNTAPLOEOLOVFETLEITGYLISAMPDLP 420
Qy 421 DLSVFQNLQVYRGRILHNGAVSLTLQIGISWLGRLSRLREGSGALILHNHTHLCFYHTV 480
    |||||
Db 421 DLSVFQNLQVYRGRILHNGAVSLTLQIGISWLGRLSRLREGSGALILHNHTHLCFYHTV 480
Qy 481 PMDQLFNNPQALLHTANRBEDECVGGLACHQICARGHCWGPPTQCVNCSQFLRQEC 540
    |||||
Db 481 PMDQLFNNPQALLHTANRBEDECVGGLACHQICARGHCWGPPTQCVNCSQFLRQEC 540
Qy 541 VEECRVLIQGLPREYVNAARHCLPCHPECOPONGSVTTCGPEPADQCVACAHYKDPFCVARC 600
    |||||
Db 541 VEECRVLIQGLPREYVNAARHCLPCHPECOPONGSVTTCGPEPADQCVACAHYKDPFCVARC 600
Qy 601 PSGVYPLSTYMPKPFDEBGAQPCPINCCHSCVYDLDDKCPAEQASPLTISVSAVVG 660
    |||||
Db 601 PSGVYPLSTYMPKPFDEBGAQPCPINCCHSCVYDLDDKCPAEQASPLTISVSAVVG 660
Qy 661 ILVVVVLGVVFGILIKRQOKIRKRYTMRRLQETELVEPLTPSGAMPNOAQMRILKETEL 720
    |||||
Db 661 ILVVVVLGVVFGILIKRQOKIRKRYTMRRLQETELVEPLTPSGAMPNOAQMRILKETEL 720
Qy 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
    |||||
Db 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVAGVSP 780
Qy 781 YVSRLLGICLTSTYQVLTQMLPYPGCLLDHRENRGRIGSODLLMWCMQIAKMSYLEDR 840
    |||||
Db 781 YVSRLLGICLTSTYQVLTQMLPYPGCLLDHRENRGRIGSODLLMWCMQIAKMSYLEDR 840
Qy 841 LVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEYHADGKVPIMMALESTLRRFT 900
    |||||

```

Db 841 LVHDDLAARNVLSVSNHVKITTDGCLARLLIDETETADGKVPDKMALLESTILRRFT 900
 Qy 901 HQSDWSYGVTVWEIMTEGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIMVKCM 960
 Db 901 HQSDWSYGVTVWEIMTEGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIMVKCM 960
 Qy 961 IDSECRPRFRELVSSEFSMADPQRFVVIYONEDLCPASPPLSTFPRSLLEDDEDDMDLVDA 1020
 Db 961 IDSECRPRFRELVSSEFSMADPQRFVVIYONEDLCPASPPLSTFPRSLLEDDEDDMDLVDA 1020
 Qy 1021 EBYLVPOGFCPPDPAPGAGGVHHRHSSSTRSGGGLTGLPESEBAPSPPLASEG 1080
 Db 1021 EBYLVPOGFCPPDPAPGAGGVHHRHSSSTRSGGGLTGLPESEBAPSPPLASEG 1080
 Qy 1081 AGSDVFDGDLQMAKGLQSLPTHDSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLQMAKGLQSLPTHDSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NQDPVRPQPSPREGCPPLPAAPACATLTERAKTSLGKNGVYKDVAFGCAVENPEYLTPO 1200
 Db 1141 NQDPVRPQPSPREGCPPLPAAPACATLTERAKTSLGKNGVYKDVAFGCAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPHPAPSPAFDNLVYMDODDPERGAPSTFKGTPAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPAPSPAFDNLVYMDODDPERGAPSTFKGTPAENPEYLGIDVPV 1255
 RESULT 4
 AAU74545 ID AAU74545 standard; Protein, 1255 AA.
 AC AAU74545;
 DT 23-APR-2002 (first entry)
 XX Human HER2 (ErbB2) polypeptide.
 DE Human HER2 (ErbB2) polypeptide.
 KM Human, HER2; ErbB; epidermal growth factor receptor; receptor;
 KM anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KM stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KM thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KM glial disorder; astrocytal disorder; hypothalamic disorder;
 KM glandular disorder; macrophagal disorder; epithelial disorder;
 KM stromal disorder; blastocoelec disorder; inflammatory disorder;
 KM angiogenic disorder; immunological disorder.
 XX Homo sapiens.
 OS
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWKOWSKI M.
 XX
 PI Erickson S, Schwall R, Slivkowski M;
 XX
 DR WPI; 2002-163686/21.
 XX
 N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor. ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 CC
 SQ Sequence 1255 AA;

Query Match 99.0%; Score 6740; DB 23; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 MELAALCRWGLLALLPGAASTOVCTGDMKLRPASPERTHLDMLRHLVQCGVYQGL 60
 Db 1 MELAALCRWGLLALLPGAASTOVCTGDMKLRPASPERTHLDMLRHLVQCGVYQGL 60
 Qy 61 ELTYLPTNASLFLQDIEVQGYVLIANNOVROVPLQRLRYRGTOLEFEDNYALAVLNG 120
 Db 61 ELTYLPTNASLFLQDIEVQGYVLIANNOVROVPLQRLRYRGTOLEFEDNYALAVLNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSTELKGVLIQRNPOLCYOTIIMKOYIRANSKF 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSTELKGVLIQRNPOLCYOTIIMKOYIRANSKF 180
 Qy 181 IGITELNRSRACHPCSPCKSGRCWGESSEDCOSLTRVYVCAAGCARCGPLPTDCHEQC 240
 Db 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCOSLTRVYVCAAGCARCGPLPTDCHEQC 240
 Qy 241 AAGCTGPRGSDCLALFHNSGICELHCPALVTYNTDFEESMPNDEGRTTFGASCVTACP 300
 Db 241 AAGCTGPRGSDCLALFHNSGICELHCPALVTYNTDFEESMPNDEGRTTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTVLCPLNNOVTAEDGTQRCBKSKPCARVYCYGIMELREVRVATSAN 360
 Db 301 YNYLSTDVGSCTVLCPLNNOVTAEDGTQRCBKSKPCARVYCYGIMELREVRVATSAN 360
 Qy 361 IQEFAGCKKIRGSLAFIPESFDGPASNTAPLOPQLOVFTLEITGYLYISAMPDSL 420
 Db 361 IQEFAGCKKIRGSLAFIPESFDGPASNTAPLOPQLOVFTLEITGYLYISAMPDSL 420
 Qy 421 DLSVFONLQVIRGRILHNHAYSLTLQIGISMLGLRSRLRELSGLALIHNTHLCFVHTV 480
 Db 421 DLSVFONLQVIRGRILHNHAYSLTLQIGISMLGLRSRLRELSGLALIHNTHLCFVHTV 480
 Qy 481 PWDOLFNNPHOALLHTANRPDECEVGBGLACHOLCARGHCWPGFTQCVNSQFLRGEC 540
 Db 481 PWDOLFNNPHOALLHTANRPDECEVGBGLACHOLCARGHCWPGFTQCVNSQFLRGEC 540
 Qy 541 VEEGRVLOGLPREVYNABHCLPCHPEOCPNGSVTCFEPEDQCVACHYVDPPEVCAR 600
 Db 541 VEEGRVLOGLPREVYNABHCLPCHPEOCPNGSVTCFEPEDQCVACHYVDPPEVCAR 600
 Qy 601 PSGVKPDLSYMPIMKFPDEBACOPCPINCHSCVDDDDKCPAEPORASPLTISYSAVVG 660
 Db 601 PSGVKPDLSYMPIMKFPDEBACOPCPINCHSCVDDDDKCPAEPORASPLTISYSAVVG 660
 Qy 661 ILVVVVGAVVFGIILKROOKIRRYTWRRLLQETELVBLPTPSGAMPNOAMRIKETEL 720
 Db 661 ILVVVVGAVVFGIILKROOKIRRYTWRRLLQETELVBLPTPSGAMPNOAMRIKETEL 720
 Qy 721 RKVAVLSGAFGTYYKGIWIPDGENVKIPVAIKYLRNTSPKAKEILDEAYVMAVGSP 780
 Db 721 RKVAVLSGAFGTYYKGIWIPDGENVKIPVAIKYLRNTSPKAKEILDEAYVMAVGSP 780
 Qy 781 YVSRLLGICLTSTVQVLTOLMPYGLDHYRENGRLGSDQLDLMWCMQIAKMSVLEDR 840

```

Db      | 781 | YVSRLLGICLTISTVQLVTLMPYGCGLDHHVRENRGLSGQDLINMCQIAKMSYLEDR 840
Qy      | 841 | LVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETEHADGGKVPKMMALESTILRRRT 900
Db      | 841 | LVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETEHADGGKVPKMMALESTILRRRT 900
Qy      | 901 | HQSDVMSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRLLQPPICTIDVYIMVKKCM 960
Db      | 901 | HQSDVMSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRLLQPPICTIDVYIMVKKCM 960
Qy      | 961 | IDSECRPRPRRELVSFSRMAPDPORFVVI QNEDLGASPLDSTFYSLLEDMDGLVDA 1020
Db      | 961 | IDSECRPRPRRELVSFSRMAPDPORFVVI QNEDLGASPLDSTFYSLLEDMDGLVDA 1020
Qy      | 1021 | EEYLVPQOGFCFCDPAFAGAGVWVHRHRSSSTRSGGDLTLGLPSEEPSPPLASPG 1080
Db      | 1021 | EEYLVPQOGFCFCDPAFAGAGVWVHRHRSSSTRSGGDLTLGLPSEEPSPPLASPG 1080
Qy      | 1081 | AGSDVFDGDLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db      | 1081 | AGSDVFDGDLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy      | 1141 | NOPDVRPOPSPRREGPLPAARAGATLEBAKTLSEKKNVVDVAFGAVENPEYLTPO 1200
Db      | 1141 | NOPDVRPOPSPRREGPLPAARAGATLEBAKTLSEKKNVVDVAFGAVENPEYLTPO 1200
Qy      | 1201 | GGAAPQHPPPAFSPADNLVYWDODPPERGA PPTSFKGTPTAENPEYLGIDVPV 1255
Db      | 1201 | GGAAPQHPPPAFSPADNLVYWDODPPERGA PPTSFKGTPTAENPEYLGIDVPV 1255

RESULT 5
AA001111
ID      | AA001111 | standard; Protein; 1255 AA.
AC      | AA001111 |
DT      | 01-JAN-1997 | (first entry)
DE      | HER-2/neu protein.
KW      | HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase; breast cancer; ovary cancer; colon cancer; lung cancer; prostate cancer; immunisation; tumour; vaccine; vector.
OS      | Homo sapiens.
FH      | Key
FT      | Domain
FT      | Location/Qualifiers
FT      | /label= "Intracellular domain
FT      | /note= "Claimed domain, useful for immunisation"

W09630514-A1.
PD      | 03-OCT-1996.
PF      | 28-MAR-1996; 96WO-US01689.
PR      | 31-MAR-1995; 95US-0414417.
PA      | (UNIW ) UNIV WASHINGTON.
PI      | Cheever MA, Disis ML;
DR      | MPI; 1996-455361/45.
DR      | N-PSDB; AAT40739.
XX      | DNA encoding HER-2-neu poly:peptide(s) - used for prevention or treatment of malignancies with which the HER-2/neu oncogene is associated
XX      | Claim 2; Page 56-61; 71pp; English.

```

```

XX      | Human HER-2/neu protein (AA001111), also called p185 or c-erbB2, is
CC      | the product of the HER-2/neu oncogene (see also AAT40739). The
CC      | protein is over-expressed in various cancers, including breast,
CC      | ovarian, colon, lung and prostate. The intracellular domain of the
CC      | protein can be used to immunise an animal against a malignancy with
CC      | which the oncogene is associated. The polypeptide can be produced
CC      | in transformed host cells for use in immunisation. Alternatively,
CC      | animal cells are transfected in vivo or ex vivo with a viral vector
CC      | that directs expression of the polypeptide.
XX      |
SQ      | Sequence 1255 AA;

```

```

Query Match      98.9%; Score 6734; DB 17; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy      | 1 | MELAALCRWGLLALLPGCASTQVCTGDMKRLRPASBETHLDMRLHYGCGVVGNI 60
Db      | 1 | MELAALCRWGLLALLPGCASTQVCTGDMKRLRPASBETHLDMRLHYGCGVVGNI 60
Qy      | 61 | ELTYLPTNASLSFLDDIQEVQGVLI IAHNQVQVPLQRLRIYRGTOLEFDNYALAVLNG 120
Db      | 61 | ELTYLPTNASLSFLDDIQEVQGVLI IAHNQVQVPLQRLRIYRGTOLEFDNYALAVLNG 120
Qy      | 121 | DPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQRNPOLCYOPTIILKQOYIRKANSKF 180
Db      | 121 | DPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQRNPOLCYOPTIILKQOYIRKANSKF 180
Qy      | 181 | IGITELNRSRACHPCSPMKSGRSGESSEDCOSLTRIVCAGGACARCKPPLPTDCHEOC 240
Db      | 181 | IGITELNRSRACHPCSPMKSGRSGESSEDCOSLTRIVCAGGACARCKPPLPTDCHEOC 240
Qy      | 241 | AAGCTGPRHSDCLACLFHNSGICELHCAALVTYNTDTRESNPBGRTRTFGASCYTAAP 300
Db      | 241 | AAGCTGPRHSDCLACLFHNSGICELHCAALVTYNTDTRESNPBGRTRTFGASCYTAAP 300
Qy      | 301 | YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREYRAVTSAN 360
Db      | 301 | YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREYRAVTSAN 360
Qy      | 361 | IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLOVFETLBEITGYLYISAMPDSL 420
Db      | 361 | IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLOVFETLBEITGYLYISAMPDSL 420
Qy      | 421 | DLVSFQNLQVIRGRILHNGAVSLTQGLGISWLGRLSRLBELSGALIHNNHLCFVHTV 480
Db      | 421 | DLVSFQNLQVIRGRILHNGAVSLTQGLGISWLGRLSRLBELSGALIHNNHLCFVHTV 480
Qy      | 481 | PWDQLFRNPQALHTANRPEDECYGEGLACHQLCARGHCMGPGPTOCVNGSOFILRGOC 540
Db      | 481 | PWDQLFRNPQALHTANRPEDECYGEGLACHQLCARGHCMGPGPTOCVNGSOFILRGOC 540
Qy      | 541 | VEECRVLQGLPREYVNAHCLPCHPECOFONGSVTCFGEADQCVAAAHYKDPFVCVAC 600
Db      | 541 | VEECRVLQGLPREYVNAHCLPCHPECOFONGSVTCFGEADQCVAAAHYKDPFVCVAC 600
Qy      | 601 | PSGVPELSTYMPIMKFPDEEGACQPCPINCHSCVDLDKGPACORASPLTSIISAVVG 660
Db      | 601 | PSGVPELSTYMPIMKFPDEEGACQPCPINCHSCVDLDKGPACORASPLTSIISAVVG 660
Qy      | 661 | ILLVVVLGVFGIILKRRQOKIRKTYMRILQETELVEPLTPSGAMPNOAQRILKETEL 720
Db      | 661 | ILLVVVLGVFGIILKRRQOKIRKTYMRILQETELVEPLTPSGAMPNOAQRILKETEL 720
Qy      | 721 | RKVKVLSSGAGCTGVYKGIWIPDGENVKI PVALKVLRENTSPYANKELIDEAVVMGVGSP 780
Db      | 721 | RKVKVLSSGAGCTGVYKGIWIPDGENVKI PVALKVLRENTSPYANKELIDEAVVMGVGSP 780
Qy      | 781 | YVSRLLGICLTISTVQLVTLMPYGCGLDHHVRENRGLSGQDLINMCQIAKMSYLEDR 840
Db      | 781 | YVSRLLGICLTISTVQLVTLMPYGCGLDHHVRENRGLSGQDLINMCQIAKMSYLEDR 840

```

QY 841 LVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETETHADGKVPPIKMMALLESILRRRT 900
DB 841 LVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETETHADGKVPPIKMMALLESILRRRT 900
QY 901 HOSDVSXGVTWELMTFGAKPYDGI PAREIPDLLEKGERLPQRPICITIDVYIMVKCM 960
DB 901 HOSDVSXGVTWELMTFGAKPYDGI PAREIPDLLEKGERLPQRPICITIDVYIMVKCM 960
QY 961 IDSECRPRRELVSSESRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMDGLVDA 1020
DB 961 IDSECRPRRELVSSESRMARDPQRFVVIQNEDLGPASPLDSTFYSLLEDDMDGLVDA 1020
QY 1021 EETLVVQGGFCFCDDPAPGACGMVHHRRSSSTRSGGDLTLGLPESEEA PRSPLAPSEG 1080
DB 1021 EETLVVQGGFCFCDDPAPGACGMVHHRRSSSTRSGGDLTLGLPESEEA PRSPLAPSEG 1080
QY 1081 ASSDVPDGDGLGMAAGLQSLPTHPDSPLORYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
DB 1081 ASSDVPDGDGLGMAAGLQSLPTHPDSPLORYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVAPPPSPREGPLPAARPAATLERAKTSLPGKNGVVKDVFAPGGAVENTEYLTTPQ 1200
DB 1141 NOPDVAPPPSPREGPLPAARPAATLERAKTSLPGKNGVVKDVFAPGGAVENTEYLTTPQ 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDOPPERGAPSPSTFKGTPTEANPEYLGLDVPIV 1255
DB 1201 GGAAPQHPPPAPSPAFDNLVYWDOPPERGAPSPSTFKGTPTEANPEYLGLDVPIV 1255

RESULT 6
AAM92406
ID AAM92406 standard; Protein; 1255 AA.
XX AAM92406;
AC 21-APR-1999 (first entry)
DT 21-APR-1999 (first entry)
XX Human HER-2/neu oncogene protein.
DE Human HER-2/neu oncogene protein.
XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
XX US5869445-A.
XX 09-FEB-1999.
XX 01-APR-1996; 96US-0625101.
XX 01-APR-1996; 96US-0625101.
XX 17-MAR-1993; 93US-0033644.
XX 12-AUG-1993; 93US-0106112.
XX 31-MAR-1995; 95US-0414417.
XX (UNIM) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX WPI; 1999-152835/13.
XX N-PSDB; AAX01912.
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
XX an HER-2/neu associated malignancy, particularly for treating or
XX preventing tumours
XX Claim 3; Column 31-38; 26pp; English.
XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 20; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPCGAASSTOVCCTDMKRLRASPETHLDMRLHYOGQOVGNL 60
DB 1 MELAALCRWGLLALLPCGAASSTOVCCTDMKRLRASPETHLDMRLHYOGQOVGNL 60
QY 61 ELTYLPTNASLSTLOIOEVQGYVLI AHNOVROPVLRRLRYVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSTLOIOEVQGYVLI AHNOVROPVLRRLRYVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTIMKDOYIKANSKF 180
DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTIMKDOYIKANSKF 180
QY 181 IGTTELNRSPACHPCSPMCKSGRCWGESSEDCSLTRTVACAGCAGCCKPLPTDCCHEQC 240
DB 181 IGTTELNRSPACHPCSPMCKSGRCWGESSEDCSLTRTVACAGCAGCCKPLPTDCCHEQC 240
QY 241 AAGCTGPKKSDCLACHFNHSGICEIHCALVTYNTDTESMNPBGRYTFGASCVTACP 300
DB 241 AAGCTGPKKSDCLACHFNHSGICEIHCALVTYNTDTESMNPBGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPHNOEVTAEEDTORCEKCSRPCARVCYGLGMEHLREYRAVTSAN 360
DB 301 YNYLSTDVSGCTLVCPHNOEVTAEEDTORCEKCSRPCARVCYGLGMEHLREYRAVTSAN 360
QY 361 IOEFACCKKIFGSLAFLPESFDGDPASNTAPLOPELOVETLEITGYLIISAMPDLP 420
DB 361 IOEFACCKKIFGSLAFLPESFDGDPASNTAPLOPELOVETLEITGYLIISAMPDLP 420
QY 421 DLSVFONLOVIRGRILHNGAYSLTLOGLISWLGRLSRLBSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFONLOVIRGRILHNGAYSLTLOGLISWLGRLSRLBSGLALIHNNTHLCFVHTV 480
QY 481 PMDQLFRNPHOALHTANRPEDECVEGGLACHOLCARGCMWGPQCVNCSQFLRGQEC 540
DB 481 PMDQLFRNPHOALHTANRPEDECVEGGLACHOLCARGCMWGPQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIWKFPEDEGACQPCPINCTHSCVDLDKGCFAEGRASPLTSIIVAVVG 660
DB 601 PSGVKPDLSTYMPIWKFPEDEGACQPCPINCTHSCVDLDKGCFAEGRASPLTSIIVAVVG 660
QY 661 ILLVVLGVVFGILIRROOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRILKETEL 720
DB 661 ILLVVLGVVFGILIRROOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS PKXNKETILDEAYMAGVSP 780
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS PKXNKETILDEAYMAGVSP 780
QY 781 YVSRILGICLTSTVOILVQLMPEYGLLDHVRENRGLSGQDLLNMCQIAKGSYLEDYR 840
DB 781 YVSRILGICLTSTVOILVQLMPEYGLLDHVRENRGLSGQDLLNMCQIAKGSYLEDYR 840
QY 841 LVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETETHADGKVPPIKMMALLESILRRRT 900
DB 841 LVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETETHADGKVPPIKMMALLESILRRRT 900


```

QY 901 HOSDWSYGVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCM 960
DB 901 HOSDWSYGVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCM 960
QY 961 ISECRPRFRELSEFSRMAIDPQRFVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1020
DB 961 ISECRPRFRELSEFSRMAIDPQRFVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1020
QY 1021 EELTVQOQGFCCDPAPGAGMWHHRSSSTRSGGDLTLGLPSESEAPRSLAPSEG 1080
DB 1021 EELTVQOQGFCCDPAPGAGMWHHRSSSTRSGGDLTLGLPSESEAPRSLAPSEG 1080
QY 1081 AGSDVDGDLGMAAKLQSLPTHDSPLQRYSEDPVPLPSETDGVAVALTCSPOPEYV 1140
DB 1081 AGSDVDGDLGMAAKLQSLPTHDSPLQRYSEDPVPLPSETDGVAVALTCSPOPEYV 1140
QY 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPO 1200
DB 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPAFNDLYYWDQDPBERGAPSTFKGPTAENPEYLGLDVPU 1255
DB 1201 GGAAPQHPPPAPSPAFNDLYYWDQDPBERGAPSTFKGPTAENPEYLGLDVPU 1255

```

RESULT 7
AAB21198 standard; protein: 1255 AA.

AAB21198;

12-JAN-2001 (first entry)

Human HER-2/neu protein.

Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
breast cancer; prostate cancer; ovarian cancer; lung cancer;

colon cancer.

Homo sapiens.

MO200044899-A1.

03-AUG-2000.

28-JAN-2000; 2000MO-US02164.

29-JAN-1999; 99US-0117976.

(COR1-) CORIXA CORP.

(SMIK) SMITHKLINE BEECHAM.

Cheever MA, Gheysen D;

WPI; 2000-505976/45.

N-PSDB; AAA89736.

HER-2/neu extracellular domain/phosphorylation domain fusion proteins
useful for vaccinating against breast, ovarian, colon, lung and
prostate cancers -

Claim 52; Fig 7; 128pp; English.

The present sequence is the human HER-2/neu protein. It is a member of
the tyrosine kinase family of receptor-like glycoproteins and shows
homology to the epidermal growth factor receptor (EGFR). It probably
plays a part in cell growth and/or differentiation. The HER-2/neu
gene is an oncogene. An HER-2/neu fusion protein comprising a
HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
domain may be used to treat or prevent cancer by eliciting or
enhancing an immune response to the HER-2/neu protein. It may be used
to treat malignancies such as breast, ovarian, colon, lung and
prostate cancers, and may be used as an antigen to vaccinate against

```

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 98.9%; Score 6734; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 MELAALCRMGILLALLPAGASTQVCTGDMKRLPASPTHLDMLRHLYQGQOVQGNL 60
DB 1 MELAALCRMGILLALLPAGASTQVCTGDMKRLPASPTHLDMLRHLYQGQOVQGNL 60
QY 61 ELTYLPTNASTSLFQDIOEVQGVLLIAHNOVRQVPLRLRIYNGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPTNASTSLFQDIOEVQGVLLIAHNOVRQVPLRLRIYNGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQSLTEILKGVLIQRNPOLCYQDTILMKDQYIKANSKF 180
DB 121 DPLNNTTPTVGTASPGGLRELQSLTEILKGVLIQRNPOLCYQDTILMKDQYIKANSKF 180
QY 181 IGITELNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVAGAGCARCKGLPTDCCHEOC 240
DB 181 IGITELNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVAGAGCARCKGLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFNHSGICELHCPALVTYNTDTFESMNPREGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFNHSGICELHCPALVTYNTDTFESMNPREGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREYAVATSAN 360
DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREYAVATSAN 360
QY 361 IOEFACCKKIFGSLAPLPSFDGDPASNTAPLOPELOQVETLEETGTGYIYAMWPSLP 420
DB 361 IOEFACCKKIFGSLAPLPSFDGDPASNTAPLOPELOQVETLEETGTGYIYAMWPSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLSLRLSGALIIHNTHLCFHYTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLSLRLSGALIIHNTHLCFHYTV 480
QY 481 PMDQLFRNPHQALLHTANRDEDECVEBGALCHQLCARGHCWGQPTQVCNCSQFLRGQEC 540
DB 481 PMDQLFRNPHQALLHTANRDEDECVEBGALCHQLCARGHCWGQPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHRCLPCHPECPQNGSVTCFEPADQCVACAHRKDPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHRCLPCHPECPQNGSVTCFEPADQCVACAHRKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDKGCFAEQRASPLTSTISAVVG 660
DB 601 PSGVKPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDKGCFAEQRASPLTSTISAVVG 660
QY 661 ILLVVLGVVFGILIKRRQOKIRKYTRRLLQSTELVEPLTPSGAMPNOQMRLKETEL 720
DB 661 ILLVVLGVVFGILIKRRQOKIRKYTRRLLQSTELVEPLTPSGAMPNOQMRLKETEL 720
QY 721 RKRYKLGSGAFGVNGIWTIPDEGNKIPAIATVLENTSPKANKELIDBAVYMAAGGSP 780
DB 721 RKRYKLGSGAFGVNGIWTIPDEGNKIPAIATVLENTSPKANKELIDBAVYMAAGGSP 780
QY 781 YVSRLLGICLTSTVOLQMPYGCLLDHYRENGRGLSGODLNMCMQIAKGSYLEDDR 840
DB 781 YVSRLLGICLTSTVOLQMPYGCLLDHYRENGRGLSGODLNMCMQIAKGSYLEDDR 840
QY 841 LVHRDLAARVNLVKSBNHYKITDFFGLARLLDIDETEVHADGKVPKIMWLESILRRRFT 900
DB 841 LVHRDLAARVNLVKSBNHYKITDFFGLARLLDIDETEVHADGKVPKIMWLESILRRRFT 900
QY 901 HOSDWSYGVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCM 960
DB 901 HOSDWSYGVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCM 960
QY 961 ISECRPRFRELSEFSRMAIDPQRFVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1020

```


Db 961 IDSCRRPREFELVSEFSRMARDPQRFVVIQNEDELGPAPPLDSTFYRSLLEDMDGLVDA 1020
Qy 1021 EEVLPVPOGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080
Db 1021 EEVLVPOGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVVAPLTGSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVVAPLTGSPQPEYV 1140
Qy 1141 NOPVRPQPSRPGSLPAPRPAATLERAKTLSPGKGVVKDVAFCGAVENPEYLTPQ 1200
Db 1141 NOPVRPQPSRPGSLPAPRPAATLERAKTLSPGKGVVKDVAFCGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPAPSPAFDNLVYWDODPERGAPSTFKGPTTAENPEYLGLDVPY 1255
Db 1201 GGAAPQHPHPAPSPAFDNLVYWDODPERGAPSTFKGPTTAENPEYLGLDVPY 1255
RESULT 8
AAV84780 ID AAV84780 standard; Protein; 1255 AA.
AC AAV84780;
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erdb-2 receptor protein.
XX
KM SPLICE erdb-2 receptor protein; cell transformation disorder; cancer;
KM tumor cell proliferation; tissue degeneration; arthropathy;
KM bone resorption; inflammatory disease; degenerative disorder;
KM wound healing.
OS Homo sapiens.
XX
XX MO200020579-A1.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99MO-CA00912.
XX
XX 02-OCT-1998; 98US-0165192.
XX
XX (UYMC-) UNIV MCMMASTER.
XX
XX Muller WJ, Siegel PM;
XX
XX WPI; 2000-303768/26.
XX
XX N-PSDB; AAA14812.
XX
XX
XX Nucleic acid encoding an erdb 2 receptor protein designated SPLICE
XX erdb-2, inhibitors of the protein are useful for treatment of cancer -
XX
XX Claim 3; Fig 2; 60pp; English.
XX
XX The present sequence represents a SPLICE erdb-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erdb-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antibodies to the
XX protein may be used to detect SPLICE erdb-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erdb-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erdb-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs, such as arthropathy, bone
XX resorption, inflammatory disease, degenerative disorders of the
XX central nervous system and wound healing.
XX
XX Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
Qy 1 MELAALCWGILLALLPPGAASVQCTGTDKRLPAPRETHLDRLHLYGCGVQVQNL 60
Db 1 MELAALCWGILLALLPPGAASVQCTGTDKRLPAPRETHLDRLHLYGCGVQVQNL 60
Qy 61 ELTYLPTVASLSFLQDIOEVQGYVLIANOVQVPLQRLVIRGTLQLEEDNALVLDNG 120
Db 61 ELTYLPTVASLSFLQDIOEVQGYVLIANOVQVPLQRLVIRGTLQLEEDNALVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSTELIKGCVLIQRNPOLCYODTILMKQYIKANSKF 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSTELIKGCVLIQRNPOLCYODTILMKQYIKANSKF 180
Qy 181 IGTIELRSPACCHCSMPCKSKSRCWGESSEDCQSITRVVCGGACRCGRLPTDCCHQC 240
Db 181 IGTIELRSPACCHCSMPCKSKSRCWGESSEDCQSITRVVCGGACRCGRLPTDCCHQC 240
Qy 241 AAGCTGPGHSDCLCLPHNHSIGICELHCPALVTYNTDFEAMPNPEGRYTGCACVTACP 300
Db 241 AAGCTGPGHSDCLCLPHNHSIGICELHCPALVTYNTDFEAMPNPEGRYTGCACVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRVATSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRVATSAN 360
Qy 361 IQEPAQCKKITGSLAFIPESFGDPASNTAPLQEBOLQVFTLEITGYLIASMPDILP 420
Db 361 IQEPAQCKKITGSLAFIPESFGDPASNTAPLQEBOLQVFTLEITGYLIASMPDILP 420
Qy 421 DLSVFQNLQVIRGRIIHNAGVSLTLOGIGISWLGRLSRLREGSGALITHNTHLCFYHTV 480
Db 421 DLSVFQNLQVIRGRIIHNAGVSLTLOGIGISWLGRLSRLREGSGALITHNTHLCFYHTV 480
Qy 481 PMDQLFRNPQALHTANRPEDECVGBGLACHOLCARGHCWPGPTQCVNCSQFLRQEC 540
Db 481 PMDQLFRNPQALHTANRPEDECVGBGLACHOLCARGHCWPGPTQCVNCSQFLRQEC 540
Qy 541 VEECRVUQGLPREVYVNAHCLPCHPECOPONGSVTCFGEPAQCAHYXDPEFCVARC 600
Db 541 VEECRVUQGLPREVYVNAHCLPCHPECOPONGSVTCFGEPAQCAHYXDPEFCVARC 600
Qy 601 PSGVKPDLSTYMPIKKFPDEBGACOPCPINCHSCVDLDDKCPAQRASPLTSIIISAVG 660
Db 601 PSGVKPDLSTYMPIKKFPDEBGACOPCPINCHSCVDLDDKCPAQRASPLTSIIISAVG 660
Qy 661 ILVVVLGVVFGILIKRQOKIRKRYTRRLQETELVEPLTPSGAMPQAOAMRILKETEL 720
Db 661 ILVVVLGVVFGILIKRQOKIRKRYTRRLQETELVEPLTPSGAMPQAOAMRILKETEL 720
Qy 721 RKRVVLGSGAFQTYKGIWIPDGENVKI PVAIKYLRNTPSKANKIILDEAYVAVAGVSP 780
Db 721 RKRVVLGSGAFQTYKGIWIPDGENVKI PVAIKYLRNTPSKANKIILDEAYVAVAGVSP 780
Qy 781 YVSRILGICLTSTVOLTQMLPYGCLLDHRENRGRIGSDLLMWCMQIAGNSYLEDVR 840
Db 781 YVSRILGICLTSTVOLTQMLPYGCLLDHRENRGRIGSDLLMWCMQIAGNSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLIDISTEYHADGAKVPIKMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLIDISTEYHADGAKVPIKMALESILRRRFT 900
Qy 901 HOSDWSYGYTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDIVYIMVXCM 960
Db 901 HOSDWSYGYTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDIVYIMVXCM 960
Qy 961 IDSCRRPREFELVSEFSRMARDPQRFVVIQNEDELGPAPPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSCRRPREFELVSEFSRMARDPQRFVVIQNEDELGPAPPLDSTFYRSLLEDMDGLVDA 1020
Qy 1021 EEVLVPOGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080

```
Db      1021 EEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSG 1080
Qy      1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOKRYSEDPVLPSETDGYVAPLTCSPQPEV 1140
Db      1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOKRYSEDPVLPSETDGYVAPLTCSPQPEV 1140
Qy      1141 NOPDVRPOPSPREGLPLAARPAAGATLEERAKTILSPKNGVNVDFAFGAVENPEYLTPO 1200
Db      1141 NOPDVRPOPSPREGLPLAARPAAGATLEERAKTILSPKNGVNVDFAFGAVENPEYLTPO 1200
Qy      1201 GGAAQPPHPPPAFAFADMLYYWDDPPERGAPESTFKGPTPAENPEYLGIDVPV 1255
Db      1201 GGAAQPPHPPPAFAFADMLYYWDDPPERGAPESTFKGPTPAENPEYLGIDVPV 1255
```

RESULT 9

AAB85458
ID AAB85458 standard; Protein; 1255 AA.

AC AAB85458;

DT 25-SEP-2001 (first entry)

DE Human HER-2/neu protein.

KM Antigen-presenting cell; immunogenic; immune response; HER-2/neu; oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

OS Homo sapiens.

PN MO200153463-A2.

PD 26-JUL-2001.

PF 19-JAN-2001; 2001WO-US01850.

PR 21-JAN-2000; 2000US-0177545.

PA (CORI-) CORIXA CORP.

PI Cheever MA, Hand-Zimmermann S;

DR WPI; 2001-476112/51.

DR N-PSDB; AAH23392.

PT New antigen-presenting cells, useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly useful for treating or preventing cancer, e.g. breast cancer

PS Claim 2; Page 41-46; 49pp; English.

CC The invention provides an isolated antigen-presenting cell, which expresses at least an immunogenic portion of a polypeptide that produces an immune response to HER-2/neu protein. The antigen-presenting cells are useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly in treating or preventing malignancies in which the HER-2/neu oncogene is associated. Specifically, these are useful for treating or preventing cancer, e.g. breast cancer, ovarian, colon, lung or prostate cancer. The present sequence represents the human HER-2/neu protein (also known as p185 or c-erbB2).

CC Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 MELALCRWGLLALIPGAASTOVCTGDMKLRILPASPEHLDMRLHYOGCQVYQGNL 60
Db      1 MELALCRWGLLALIPGAASTOVCTGDMKLRILPASPEHLDMRLHYOGCQVYQGNL 60
Qy      61 ELTYLPTNASTLSFLDIOEVQGVYLIAHQVNRQVPLQRLIRIVGTQLFEDNYALAVLDNG 120
```

```
Db      61 ELTYLPTNASTLSFLDIOEVQGVYLIAHQVNRQVPLQRLIRIVGTQLFEDNYALAVLDNG 120
Qy      121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLIORNPQLCYQDTLLKQDYIKANSKF 180
Db      121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLIORNPQLCYQDTLLKQDYIKANSKF 180
Qy      181 IGTILNRSRACHPCSPMKCKSRGCESESDQSILRTYTCAGGCARCKPLPTDCHEOC 240
Db      181 IGTILNRSRACHPCSPMKCKSRGCESESDQSILRTYTCAGGCARCKPLPTDCHEOC 240
Qy      241 AAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTEESMPNPEGRYTFGASCTYAC 300
Db      241 AAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTEESMPNPEGRYTFGASCTYAC 300
Qy      301 VNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLAMEHLREYRAVTSAN 360
Db      301 VNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLAMEHLREYRAVTSAN 360
Qy      361 IOEFAGCKKI FGSILAFLESFDDGDPASNTAPLOPELOVFEETLEETGYLYISAMPDSL 420
Db      361 IOEFAGCKKI FGSILAFLESFDDGDPASNTAPLOPELOVFEETLEETGYLYISAMPDSL 420
Qy      421 DLSVFQNLQVIRGRIILHNGAVSLTLQGLGISWLAGLSRLRELSGLALIHNNHLCFVHTV 480
Db      421 DLSVFQNLQVIRGRIILHNGAVSLTLQGLGISWLAGLSRLRELSGLALIHNNHLCFVHTV 480
Qy      481 PMDQLFRNHQALLHTANRPEDECVGEGLACHQLCARHGWGPGPTQCVNCSQFLRGQEC 540
Db      481 PMDQLFRNHQALLHTANRPEDECVGEGLACHQLCARHGWGPGPTQCVNCSQFLRGQEC 540
Qy      541 VEECVTLQGLPREVYNARHCLPCHPECOPONSVCYCFGEADQCVACAHYKOPFCVAC 600
Db      541 VEECVTLQGLPREVYNARHCLPCHPECOPONSVCYCFGEADQCVACAHYKOPFCVAC 600
Qy      601 PSGVAPDLSYMPIMKFPDEEGACQCPINCTHSCVDLDDKGCBAEORASPLTSIVAVVG 660
Db      601 PSGVAPDLSYMPIMKFPDEEGACQCPINCTHSCVDLDDKGCBAEORASPLTSIVAVVG 660
Qy      661 ILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOAKRILKETEL 720
Db      661 ILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOAKRILKETEL 720
Qy      721 RKVKVLGSGAGCTVYKGIPIPDGENVKIPVAIKVLRENTSPKANETILDEAVYMAVGSP 780
Db      721 RKVKVLGSGAGCTVYKGIPIPDGENVKIPVAIKVLRENTSPKANETILDEAVYMAVGSP 780
Qy      781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVENRGLSGQDLNMCQIAKGSYLEDR 840
Db      781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVENRGLSGQDLNMCQIAKGSYLEDR 840
Qy      841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETERYHADGKVPKIMALESIILRRFT 900
Db      841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETERYHADGKVPKIMALESIILRRFT 900
Qy      901 HOSDWSVGVYWEMLTREGAKKEYDGI PAEIRIDILEKGRLLQPPCTIDVYIMYKCM 960
Db      901 HOSDWSVGVYWEMLTREGAKKEYDGI PAEIRIDILEKGRLLQPPCTIDVYIMYKCM 960
Qy      961 IDSECRPFRELVSFSRMAARDPQRFVYIQNEDLPASPLDSTFRSLLEDMDMDLYA 1020
Db      961 IDSECRPFRELVSFSRMAARDPQRFVYIQNEDLPASPLDSTFRSLLEDMDMDLYA 1020
Qy      1021 EEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSG 1080
Db      1021 EEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSG 1080
Qy      1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOKRYSEDPVLPSETDGYVAPLTCSPQPEV 1140
Db      1081 AGSDVFDGDLGMAAKGLOSLPTHDPSPLOKRYSEDPVLPSETDGYVAPLTCSPQPEV 1140
Qy      1141 NOPDVRPOPSPREGLPLAARPAAGATLEERAKTILSPKNGVNVDFAFGAVENPEYLTPO 1200
```

Db 1141 NOPDVRPDPSPBEGPLPAAPAGATLERPKTSLSPKNGVKVDFAFGAVENPEYLTPQ 1200
QY 1201 GGAAPQHPHPAFSPAFDNLYYWDDPPERGAPSPSTFKGTPTAENPEYLGLDVPV 1255
DB 1201 GGAAPQHPHPAFSPAFDNLYYWDDPPERGAPSPSTFKGTPTAENPEYLGLDVPV 1255
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000MO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIIM-) EPIIMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E,
PI Keogh E;
XX
DR WP1; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
PS Disclosure; Page 15; 1999p; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (II), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;
Query Match 98.9%; Score 6734; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 METALCRWGLLALPPAASTOYCTGDMTLRLPASFTLMDMLRHLYXCCGVVQGNL 60
DB 1 METALCRWGLLALPPAASTOYCTGDMTLRLPASFTLMDMLRHLYXCCGVVQGNL 60
QY 61 ELTYLPTNASLSFLDIOEVQGVYLIAHQVQVPLQRLIRYGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYLIAHQVQVPLQRLIRYGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELOLRSLTEILKGVLLIQNPOLCYODTILMKDYIKANSKF 180
DB 121 DPLNNTTPTVGASPGGLRELOLRSLTEILKGVLLIQNPOLCYODTILMKDI FKHNNQLA 180
QY 181 IGITELNSRACHPSSPCCKGRCGSESEDDQSILTRYCAGCCARCKPPLTDCCHEC 240
DB 181 LTLIDTNSRACHPSSPCCKGRCGSESEDDQSILTRYCAGCCARCKPPLTDCCHEC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDPPESMPNEGRTTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDPPESMPNEGRTTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLYVCPLEHNOEVTABDGTORCEKSPCARVCYGLMEHLREVAVTSAN 360
DB 301 YNYLSTDVGSCTLYVCPLEHNOEVTABDGTORCEKSPCARVCYGLMEHLREVAVTSAN 360
QY 361 IOEFAGCKKIRGSLAFIPESPDPGDPASNTAPLQEPOLQVFELEITGLYISAMPDSLP 420
DB 361 IOEFAGCKKIRGSLAFIPESPDPGDPASNTAPLQEPOLQVFELEITGLYISAMPDSLP 420
QY 421 DLSVFQNTQVLRGRILHNGAVSLTLQGLISMLGLRSLRELGSLALHHNTHLCFVHTV 480
DB 421 DLSVFQNTQVLRGRILHNGAVSLTLQGLISMLGLRSLRELGSLALHHNTHLCFVHTV 480
QY 481 PWDOLFERNPHOALLTANRPEDECVGEGILACGHCARHCMPGPTQVCNCSOFLRGQEC 540
DB 481 PWDOLFERNPHOALLTANRPEDECVGEGILACGHCARHCMPGPTQVCNCSOFLRGQEC 540
QY 541 VEECVVLQGLPREYVNAHCLPCHPECOPONGSVTCFGBADQCVACAHYKDPFCVAPC 600
DB 541 VEECVVLQGLPREYVNAHCLPCHPECOPONGSVTCFGBADQCVACAHYKDPFCVAPC 600
QY 601 PSGVPELDSYPIWPFPEBEGACOPCPINCHSCVDLDDKGPABORASPLTSYSAVVG 660
DB 601 PSGVPELDSYPIWPFPEBEGACOPCPINCHSCVDLDDKGPABORASPLTSYSAVVG 660
QY 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQOQMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQOQMRILKETEL 720
QY 721 RKVKYLSGAGTGYKGIWIDGENVKI PVAIKVIRENTSPRANKEILDEAVYMGVSP 780
DB 721 RKVKYLSGAGTGYKGIWIDGENVKI PVAIKVIRENTSPRANKEILDEAVYMGVSP 780
QY 781 YVSRLLGICLSTYQVLVQLMPYGLDLHVENRRLSODLNNCMOIAKMSYLEVDV 840
DB 781 YVSRLLGICLSTYQVLVQLMPYGLDLHVENRRLSODLNNCMOIAKMSYLEVDV 840
QY 841 LVHRDLAARNLVKSPNHTKITDGLARLIDETEHYADGKVPYIKMMALESIARRFT 900
DB 841 LVHRDLAARNLVKSPNHTKITDGLARLIDETEHYADGKVPYIKMMALESIARRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVCKM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVCKM 960
QY 961 IDSECRPFRLVSEFSMARDDPQRFVYIQNEDJGAPASPLDSTFRSLLEDMDGDLYDA 1020
DB 961 IDSECRPFRLVSEFSMARDDPQRFVYIQNEDJGAPASPLDSTFRSLLEDMDGDLYDA 1020
QY 1021 EBYLVPOGFPCCPDPAAGGMVHHRHSSSTRSGGDLTLGLSPSEEARSPPLAPSEG 1080
DB 1021 EBYLVPOGFPCCPDPAAGGMVHHRHSSSTRSGGDLTLGLSPSEEARSPPLAPSEG 1080

Dh 1021 EEYLVPQGGFCPPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQVSEDPVPLPSETDGYVAPLTCSPQPEVY 1140
Dh 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQVSEDPVPLPSETDGYVAPLTCSPQPEVY 1140
Qy 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200
Dh 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200
Qy 1201 GGAAPOPHPPAPSPAFNDLYYWDODPPERGAPSTFKGPTTANENPEYLGADVPV 1255
Dh 1201 GGAAPOPHPPAPSPAFNDLYYWDODPPERGAPSTFKGPTTANENPEYLGADVPV 1255

RESULT 11

AAE24067

ID AAE24067 standard; Protein; 1255 AA.

AAE24067;

DT 23-SEP-2002 (first entry)

DE Human Her-2 protein.

KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.

OS Homo sapiens.

PN WO200222636-A1.

PD 21-MAR-2002.

PF 12-SEP-2001; 2001WO-US28572.

PR 15-SEP-2000; 2000US-0663834.

PA (ISIS-) ISIS PHARM INC.

PI Bennett CF, Coweert LM;

DR WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT Inflammation or to prevent infection in humans -

PS Example 13; Page 95-107; 116pp; English.

CC The invention relates to antisense compounds targeted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

SQ Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MEIALALCRWGLLALALPGAASTOVCTGTDMLRLPASPEHLMDLRLHYOGQVQGNL 60
Dh 1 MEIALALCRWGLLALALPGAASTOVCTGTDMLRLPASPEHLMDLRLHYOGQVQGNL 60

Qy 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200

Qy 61 ELTYLPTNALSFLQDIOEVQGVLLIAHQVROVPLQRLRIVRGTOLFEDNVALAVALDNG 120
Dh 61 ELTYLPTNALSFLQDIOEVQGVLLIAHQVROVPLQRLRIVRGTOLFEDNVALAVALDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLIQRNPOLCYODTILMKDOYIKANSKF 180
Dh 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLIQRNPOLCYODTILMKDOYIKANSKF 180
Qy 181 IGITELNRACHPCSPMKSGRCSGSESEDDOSLTRTCAGGCARCKPLPDCCHEOC 240
Dh 181 LTLIDNRSRACHPCSPMKSGRCSGSESEDDOSLTRTCAGGCARCKPLPDCCHEOC 240
Qy 241 AAGCTPKRSDCLACLFHNSGICELHCALVTYNTDTESMNPREGRYFFGASCVTACP 300
Dh 241 AAGCTPKRSDCLACLFHNSGICELHCALVTYNTDTESMNPREGRYFFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSPKARVCYGLCMEHLREVRAVTSAN 360
Dh 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSPKARVCYGLCMEHLREVRAVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFPLPESPDGPASNTAPLQEQLOVFETLEEITGYLYISAMPDSL 420
Dh 361 IOEFAGCKKIFGSLAFPLPESPDGPASNTAPLQEQLOVFETLEEITGYLYISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRIIHNAGVSLTLQGLGISWLGRLSRLSGSLALIHNTILCFVHTV 480
Dh 421 DLSVFQNLQVIRGRIIHNAGVSLTLQGLGISWLGRLSRLSGSLALIHNTILCFVHTV 480
Qy 481 PMDQLFRNPQALHTANRPEDECVGEGLAGHQLCARHCWGPPTQVCNCSQFLRGQEC 540
Dh 481 PMDQLFRNPQALHTANRPEDECVGEGLAGHQLCARHCWGPPTQVCNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECOQONSYYTCFGEBAOCVACAHKDPFCVARC 600
Dh 541 VEECRVLOGLPREYVNAHCLPCHPECOQONSYYTCFGEBAOCVACAHKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIMKPEDEGACQPCPINCTHSCVDLDKGCFAEGRASPLTISVAVNG 660
Dh 601 PSGVKPDLSPYMPIMKPEDEGACQPCPINCTHSCVDLDKGCFAEGRASPLTISVAVNG 660
Qy 661 ILLVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOAKMILKETEL 720
Dh 661 ILLVVVLGVVFGILLIKRQOKIRKTYMRRLQETELVEPLTPSGAMPNOAKMILKETEL 720
Qy 721 RKVKVLSGAFGVVKGIMI PGGENVKIPVAIKVRENTSPKANKETILDEAYVMAVGSP 780
Dh 721 RKVKVLSGAFGVVKGIMI PGGENVKIPVAIKVRENTSPKANKETILDEAYVMAVGSP 780
Qy 781 YVSRLLGICLTSTVOLVTLQMLPYGCLLDHVRNRRGLSGODLLNMCQIAKGSYLEYDR 840
Dh 781 YVSRLLGICLTSTVOLVTLQMLPYGCLLDHVRNRRGLSGODLLNMCQIAKGSYLEYDR 840
Qy 841 LVHRDLAANVVLKSPBNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESILRRFT 900
Dh 841 LVHRDLAANVVLKSPBNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESILRRFT 900
Qy 901 HOSDWSYGVVWELMTBEAKKYDGI PAEIPDLLEKGRLLQPPCTIDVYIMYKCMW 960
Dh 901 HOSDWSYGVVWELMTBEAKKYDGI PAEIPDLLEKGRLLQPPCTIDVYIMYKCMW 960
Qy 961 IDSECRPRRELVSERSMARDPQRFVYQNEIDLGPASPLDSTFYSLLEDDMGDLVDA 1020
Dh 961 IDSECRPRRELVSERSMARDPQRFVYQNEIDLGPASPLDSTFYSLLEDDMGDLVDA 1020
Qy 1021 EEYLVPQGGFCPPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Dh 1021 EEYLVPQGGFCPPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQVSEDPVPLPSETDGYVAPLTCSPQPEVY 1140
Dh 1081 AASDVFDGDLGMAAGKGLSLPTHDPSPLOQVSEDPVPLPSETDGYVAPLTCSPQPEVY 1140
Qy 1141 NOPDVRPOPSPREGEPLPAARPAAGATLERAKTSLPGKNGVXKDVAFGGAVENTEYLTPQ 1200

Db 1141 NOPDRPDPSPREBPPLPAARPAATLEPKTLSPKNGVAVDVAFGAVNPEYLPO 1200
Qy 1201 GGAAPQHPHPAPSPAFDNLYYMDODPEERGAPESTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPAPSPAFDNLYYMDODPEERGAPESTFKGTPTAENPEYLGIDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
XX
XX
XX 01-JUL-2002 (first entry)
XX
XX Human Her-2/neu protein.
XX
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 1021..1030
XX FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
XX WO200214503-A2.
XX
XX 21-FEB-2002.
XX
XX 14-AUG-2001; 2001WO-US41733.
XX
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-236428P.
XX 21-FEB-2001; 2001US-270520P.
XX
XX (CORI-) CORIYA CORP.
XX
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
XX Mcneill PD, Vedvick TS;
XX
XX WPI: 2002-280758/32.
XX DR N-PSDB; AAD32743.
XX
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX prevention and diagnosis of cancer, preferably breast cancer
XX
XX
XX Disclosure; Page 114-117; 129pp; English.
XX
XX The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX The composition is useful for the therapy and diagnosis of cancer,
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX and other compositions for the diagnosis, prevention and treatment of
XX human malignancies, for stimulating and/or expanding T cells specific for
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX patient. The invention is useful for stimulating a T cell response in a
XX human patient, as probe or primer for nucleic acid hybridisation, to
XX selectively form duplex molecules with complementary stretches of the
XX entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX length gene from a suitable library, and to direct expression of a
XX polypeptide in appropriate host cells. The composition is useful in
XX prophylactic or therapeutic applications and for the treatment of cancer,
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
XX
XX Sequence 1255 AA;
Query Match 98.9%; Score 6734; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
Qy 1 METALCGMGLIALLPGAASSTOYCTGDMTLRLPASETHLDMRLHYOCGVVQGL 60
Db 1 METALCGMGLIALLPGAASSTOYCTGDMTLRLPASETHLDMRLHYOCGVVQGL 60
Qy 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVROVLORLIRVGTOLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVROVLORLIRVGTOLFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELOLRSLTEILKGVLIQRNPOLCYDITLMKDYIKANSKF 180
Db 121 DPLNNTTPTVGTASPGGLRELOLRSLTEILKGVLIQRNPOLCYDITLMKDYIKANSKF 180
Qy 181 IGITELNRSRACHPCSPCKGSRCKGSESEDOSLTRVVCAGGACARCGPLTDCHECC 240
Db 181 LTLIDTNSRCHPCSPCKGSRCKGSESEDOSLTRVVCAGGACARCGPLTDCHECC 240
Qy 241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCYTACP 300
Db 241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCYTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNNOEVTADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNNOEVTADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEPAGCKKIRGSLAFILPESFDGDPASNTAPLOPEOLOVFETLEITGYLI SANPDSL 420
Db 361 IOEPAGCKKIRGSLAFILPESFDGDPASNTAPLOPEOLOVFETLEITGYLI SANPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELGSGALILHNNHLCFVHV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELGSGALILHNNHLCFVHV 480
Qy 481 PMDOLFRNPQALHTANRPEDECVEGLACHOLCARGHCGPPTCCVNCQSFRLRGDEC 540
Db 481 PMDOLFRNPQALHTANRPEDECVEGLACHOLCARGHCGPPTCCVNCQSFRLRGDEC 540
Qy 541 VEECRVLOGLPREVYNARHCIPCHPECCPONGSYTCFPEPADOCYACAHYKDPFCVARC 600
Db 541 VEECRVLOGLPREVYNARHCIPCHPECCPONGSYTCFPEPADOCYACAHYKDPFCVARC 600
Qy 601 PSGVCPDLSYMPIKFPDEBGACOPCPINCHSCCYDLDDKCPAQRASPLTISYAVVG 660
Db 601 PSGVCPDLSYMPIKFPDEBGACOPCPINCHSCCYDLDDKCPAQRASPLTISYAVVG 660
Qy 661 ILLVVLGVVEFGILIKRQOKIRRYTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVEFGILIKRQOKIRRYTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVYVLGGAGTYKKGWIPDGENVKIPVAKYLRRENTSPRANKELIDEAYVWAGVSP 780
Db 721 RKVYVLGGAGTYKKGWIPDGENVKIPVAKYLRRENTSPRANKELIDEAYVWAGVSP 780
Qy 781 YVSRLLGICLSTVQVLTQMLPYGCLDHYENNGRLSGDOLLMQMOIAGMSYLEVVR 840
Db 781 YVSRLLGICLSTVQVLTQMLPYGCLDHYENNGRLSGDOLLMQMOIAGMSYLEVVR 840
Qy 841 LVHRLDAAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKYPKIMMALESILRRFT 900
Db 841 LVHRLDAAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKYPKIMMALESILRRFT 900
Qy 901 HQSVNVSQVTVWMLTFGAKPYGIGIPARELPDLLEGERLPQPICITIVYIMVWCM 960
Db 901 HQSVNVSQVTVWMLTFGAKPYGIGIPARELPDLLEGERLPQPICITIVYIMVWCM 960
Qy 961 IDSECRPRELVESEFRMARDDPORFVYIWNEDIGPASPLDSTYRSLLEDMDGDLVDA 1020
Db 961 IDSECRPRELVESEFRMARDDPORFVYIWNEDIGPASPLDSTYRSLLEDMDGDLVDA 1020
Qy 1021 EBYLVPQGGFFCPDPAGAGMVHRRSSSTSGGDLTLGLPSESEAPRSLAPASEG 1080
Db 1021 EBYLVPQGGFFCPDPAGAGMVHRRSSSTSGGDLTLGLPSESEAPRSLAPASEG 1080

Db 1021 EBYLVPOQGFPCDPAPGAGWVHRHRSSSTRSGCGDLTLGLPSEBEPAPRSPPLAPSBG 1080
QY AGSDVPDDDLGGAAGKGLQSLPTHPSPLQRYSEDPVPLPSETDGYAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGGAAGKGLQSLPTHPSPLQRYSEDPVPLPSETDGYAPLTCSPQPEYV 1140
QY 1141 NOPDVRPOPSPREBEPPLAARPAAGATLERAKTSLPGKNGVADVAFAGAVENPEYLTPO 1200
Db 1141 NOPDVRPOPSPREBEPPLAARPAAGATLERAKTSLPGKNGVADVAFAGAVENPEYLTPO 1200
QY 1201 GGAAPOPHPPAPFSPAFDNLVYWDODPPERGAPESTFKCTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPOPHPPAPFSPAFDNLVYWDODPPERGAPESTFKCTPTAENPEYLGIDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

DT 17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

KM Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX tyrosine kinase; receptor; c-erbB2; gene therapy.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Domain 1..653 /note="extracellular domain"

FT Domain 676..1255 /note="intracellular domain"

FT Domain 990..1255 /note="phosphorylation domain"

PN WO200212341-A2.

PD 14-FEB-2002.

PF 03-AUG-2001; 2001WO-US24283.

PR 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Cheever MA, Gheysen D;

DR WPI; 2002-241743/29.

DR N-PsDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

PT or enhancing an immune response to the protein, has Her-2/neu

PT extracellular domain fused to Her-2/neu intracellular or

PT phosphorylation domain -

PS Claim 68; Fig 7; 14pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaCD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal *ex vivo* with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

SQ Sequence 1255 AA;

Query Match 98.9%; Score 6734; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELALCWGILLALLPGGAASVYCTGTDKTLRLPASPETHLDMRLHYOGCQVVGSL 60
Db 1 MELALCWGILLALLPGGAASVYCTGTDKTLRLPASPETHLDMRLHYOGCQVVGSL 60
QY 61 ETTYLPTNASISFLDDIOEVQGYVLIANNQVQVLOLRIRYGTQLFEDNALAVLDNG 120
Db 61 ETTYLPTNASISFLDDIOEVQGYVLIANNQVQVLOLRIRYGTQLFEDNALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDQYIRANSKF 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDQYIRANSKF 180
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDQYIRANSKF 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDQYIRANSKF 180
QY 181 IGITELNNSRACHPESPMCKSKSRCESSSDQSLTRYVACAGCARCKPLTDCHEQC 240
Db 181 IGITELNNSRACHPESPMCKSKSRCESSSDQSLTRYVACAGCARCKPLTDCHEQC 240
QY 241 AAGCTGPRHSDCLACLNHNSGICELHCPALVYTTDTFESMPNEGRTTFASCYTACP 300
Db 241 AAGCTGPRHSDCLACLNHNSGICELHCPALVYTTDTFESMPNEGRTTFASCYTACP 300
QY 301 YNYLSTDVGSCTLVCPPLNQEVTAEDGTORCEKSKPCARVCYGLMEHLREAVATSAN 360
Db 301 YNYLSTDVGSCTLVCPPLNQEVTAEDGTORCEKSKPCARVCYGLMEHLREAVATSAN 360
QY 361 IOEPAGCKKIRGSLAFPESPDGPASNTAPLPQPOLQVFTLEITGLYLSAMPDSLP 420
Db 361 IOEPAGCKKIRGSLAFPESPDGPASNTAPLPQPOLQVFTLEITGLYLSAMPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNHGAVALTLQGLGISMGLRSRLRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNHGAVALTLQGLGISMGLRSRLRELGSGLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPQALHTANRPEDECVBEGLAGHQLCARHGWGPGTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPQALHTANRPEDECVBEGLAGHQLCARHGWGPGTQCVNCSQFLRGQEC 540
QY 541 VEBECVLOGLPREYVNAHCLPCHPEQCQPNQSVTCFGBEADQCAACHYKDPPEVCARC 600
Db 541 VEBECVLOGLPREYVNAHCLPCHPEQCQPNQSVTCFGBEADQCAACHYKDPPEVCARC 600
QY 601 PSGVPRDLSYMPIMKFPBEGACQPCPINCTHSCVLDLDDKGPAPORASPLTSISAVVG 660
Db 601 PSGVPRDLSYMPIMKFPBEGACQPCPINCTHSCVLDLDDKGPAPORASPLTSISAVVG 660
QY 661 ILLVVVIGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOQRILKETEL 720
Db 661 ILLVVVIGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOQRILKETEL 720
QY 721 RKVKVLSSGAGCTYVKGWIWDGENVKIPVAIKYLRNTSPRANKETIDEAVMAGVQSP 780
Db 721 RKVKVLSSGAGCTYVKGWIWDGENVKIPVAIKYLRNTSPRANKETIDEAVMAGVQSP 780


```

QY 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNENRGLSGSODLLNMCQIAKMSYLEDR 840
DB 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNENRGLSGSODLLNMCQIAKMSYLEDR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETRYHADGGKVPKIMMALESILRRRT 900
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETRYHADGGKVPKIMMALESILRRRT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRPLPQPICTIDVYMIMVKCM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRPLPQPICTIDVYMIMVKCM 960
QY 961 IDSECRPRRELVSERSRMARDPQRFVVIQNEIDLGPASLDSTFYNSLLEDDMGDLVDA 1020
DB 961 IDSECRPRRELVSERSRMARDPQRFVVIQNEIDLGPASLDSTFYNSLLEDDMGDLVDA 1020
QY 1021 EBYLVPOOGFCPPDAPAGAGVWHRHRSSTRSGGDLTLGLEPSEERAPSPAPSPSG 1080
DB 1021 EBYLVPOOGFCPPDAPAGAGVWHRHRSSTRSGGDLTLGLEPSEERAPSPAPSPSG 1080
QY 1081 AGSDVPDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVPDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPSRREGPLPAARAGATLEBAKTLSEKKGAVKQVAFGSAVENPEYLPQ 1200
DB 1141 NOPDVRPQPSRREGPLPAARAGATLEBAKTLSEKKGAVKQVAFGSAVENPEYLPQ 1200
QY 1201 GGAAPQHPPPAFSAFDNLYYMDOPPERGAPSPSTFKGTPTAENPEYLGIDVPV 1255
DB 1201 GGAAPQHPPPAFSAFDNLYYMDOPPERGAPSPSTFKGTPTAENPEYLGIDVPV 1255

```

RESULT 14
AAU77114 standard; Protein; 1255 AA.

```

AC AAU77114;
DT 05-JUN-2002 (first entry)
DE Human Her-2/neu polypeptide.
XX Human, Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
OS Homo sapiens.
XX W0200213847-A2.
XX 21-FEB-2002.
XX 13-AUG-2001; 2001WO-US25408.
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX (CORI-) CORIXA CORP.
XX
XX Gaigier A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
XX encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX

```

```

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA:
Query Match 98.9%; Score 6734; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 METALACRWGLLALLPGAASTQVCTGDMKLRPASETHLDMRLHYOGQVQGNL 60
DB 1 METALACRWGLLALLPGAASTQVCTGDMKLRPASETHLDMRLHYOGQVQGNL 60
QY 61 ELTYLPTNASLSEFLDIOEVQGVLIANOVQVQLQRLIRYRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSEFLDIOEVQGVLIANOVQVQLQRLIRYRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPQLCYOPTILWKDOYIKANSXF 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPQLCYOPTILWKDOYIKANSXF 180
QY 181 IGITELNRSRACHPCSPCKSGRCSESSSDQSLTRTVACAGCARCKGRLPTDCCHEQC 240
DB 181 IGITELNRSRACHPCSPCKSGRCSESSSDQSLTRTVACAGCARCKGRLPTDCCHEQC 240
QY 241 AAGCTGPRGSDCLACGHNHSGICELHCPALVTVTTDTFESHNPNEGRTTFCASCYTACP 300
DB 241 AAGCTGPRGSDCLACGHNHSGICELHCPALVTVTTDTFESHNPNEGRTTFCASCYTACP 300
QY 301 YNYLSTDVGSCTVCPPLHNOEVTABEDGTORCEKSGPCARVCYGGMEHLREVRVAVTSAN 360
DB 301 YNYLSTDVGSCTVCPPLHNOEVTABEDGTORCEKSGPCARVCYGGMEHLREVRVAVTSAN 360
QY 361 IOEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPBOLQVFBTEIEITGLYISAMPDSL 420
DB 361 IOEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPBOLQVFBTEIEITGLYISAMPDSL 420
QY 421 DLSVFNQVIRGRILHNGAVSLTLQGGISLWGLRSRLRELSGSLALIHNTHLCFVHTV 480
DB 421 DLSVFNQVIRGRILHNGAVSLTLQGGISLWGLRSRLRELSGSLALIHNTHLCFVHTV 480
QY 481 PMDQLFRNPQALHTANRPEDECYEGELACHOLCARGHCGPPTOCVNCQSQFLRGQEC 540
DB 481 PMDQLFRNPQALHTANRPEDECYEGELACHOLCARGHCGPPTOCVNCQSQFLRGQEC 540
QY 541 VEECRVLQGLREYVNAHCLPCHPBCOPONGSVTCFPEADQCVACHYADPPFCVARC 600
DB 541 VEECRVLQGLREYVNAHCLPCHPBCOPONGSVTCFPEADQCVACHYADPPFCVARC 600
QY 601 PSQVPRDLSVPIKPFDEBGAQPCPINTCHSCVDLDDKCCPAQORASPLTISIISAVYG 660
DB 601 PSQVPRDLSVPIKPFDEBGAQPCPINTCHSCVDLDDKCCPAQORASPLTISIISAVYG 660
QY 661 ILVVVVGAVVFGILIKRQOKIRKYTRRLQETELVEPLTPSGAMPQAOAMRLIKETEL 720
DB 661 ILVVVVGAVVFGILIKRQOKIRKYTRRLQETELVEPLTPSGAMPQAOAMRLIKETEL 720
QY 721 RKVAVLSSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVWAGVSP 780
DB 721 RKVAVLSSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVWAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNENRGLSGSODLLNMCQIAKMSYLEDR 840
DB 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNENRGLSGSODLLNMCQIAKMSYLEDR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETRYHADGGKVPKIMMALESILRRRT 900

```

DB 841 LVHRLAARNVLYKSPNHVKITDFGLARLDDIDETEHADGSKVPIKMALESILRRRT 900
QY 901 HOSDWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960
DB 901 HOSDWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960
QY 961 IDSECRPRRELVSFESPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLIEDDMGDLYNA 1020
DB 961 IDSECRPRRELVSFESPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLIEDDMGDLYNA 1020
QY 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPTHDSPLOQRYSEDPVPLPSETDGYVAPLTCSPOPEYV 1140
DB 1081 AGSDVFDGDLGMAKGLQSLPTHDSPLOQRYSEDPVPLPSETDGYVAPLTCSPOPEYV 1140
QY 1141 NOPDVRPOPSPREGLPARPAGATLEPAKTLSPGKGVVMDVFAFGAVENPEYLTPO 1200
DB 1141 NOPDVRPOPSPREGLPARPAGATLEPAKTLSPGKGVVMDVFAFGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPAPFAFDNLVYWDOPPERGAPRSTFKGTPTAENPEYLGLDVPV 1255
DB 1201 GGAAPQHPHPAPFAFDNLVYWDOPPERGAPRSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
XX
XX 07-FEB-1994 (first entry)
XX
XX Sequence of c-erbB-2 tumour antigen.
XX
XX Tumour antigen: c-erbB-2; glycoprotein.
XX
XX Homo sapiens.
XX
XX WO9316185-A.
XX
XX PD 19-AUG-1993.
XX
XX PF 05-FEB-1993; 93WO-US01055.
XX
XX PR 06-FEB-1992; 92US-0831967.
XX
XX PA (CEFTU) CEFTUS ONCOLOGY CORP.
XX PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
XX Houston LT, Huston JS, Oppermann H, Ring DB;
XX
XX WPI: 1993-272889/34.
XX
XX DR N-PSDB; AAQ46083.
XX
XX PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX
XX PS Disclosure; pages 48-54; 87pp; English.
XX
XX CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells. Such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
XX the location of a stop codon in AAQ46083.
XX
XX Sequence 1433 AA;

Query Match 98.3%; Score 6691; DB 14; Length 1433;
Best Local Similarity 98.2%; Pred. No. 0;

Matches 1233; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
QY 1 MELALCWMGLLALLPFGAASVCTGTDMKTLRIPASPTLIDMLRHLVQSCVQVQNL 60
DB 1 MELALCWMGLLALLPFGAASVCTGTDMKTLRIPASPTLIDMLRHLVQSCVQVQNL 60
QY 61 ELTYLPTNASLSFLDIOEVQGYVLIANNQVQVLPQRIYRGTLQFEDNALAVLNDG 120
DB 61 ELTYLPTNASLSFLDIOEVQGYVLIANNQVQVLPQRIYRGTLQFEDNALAVLNDG 120
QY 121 DPLANTTPVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDQYIKANSKF 180
DB 121 DPLANTTPVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDQYIKANSKF 180
QY 181 IGITELNRSRACHPCSPCKSGRCGSESEDCOSLTRVYACGACRCARCPRTDCHEOC 240
DB 181 IGITELNRSRACHPCSPCKSGRCGSESEDCOSLTRVYACGACRCARCPRTDCHEOC 240
QY 241 AAGCTGPRKNSDCLAFHNSGICELHCPALVTYNTDPESPNPEGRTFGASCVTACP 300
DB 241 AAGCTGPRKNSDCLAFHNSGICELHCPALVTYNTDPESPNPEGRTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPRLHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVATSN 360
DB 301 YNYLSTDVGSCTLVCPRLHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVATSN 360
QY 361 IOEPAGCKKIFGSLAFPLESPFDGPASNTAPLOPEQLOVFEYLEITGYLYISAMPDSL 420
DB 361 IOEPAGCKKIFGSLAFPLESPFDGPASNTAPLOPEQLOVFEYLEITGYLYISAMPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAVSLTLOGISWLGRLSRLREGSGALIHNNHLCVHV 480
DB 421 DLSVFQNLQVIRGRILHNGAVSLTLOGISWLGRLSRLREGSGALIHNNHLCVHV 480
QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGEC 540
DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGEC 540
QY 541 VEECVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAACAHYKDPFCVARC 600
DB 541 VEECVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAACAHYKDPFCVARC 600
QY 601 PSGVPRDLSYMPIMKFPDEBEGACOPCINCHSCVDLDKGPAGORASPLTSYSAVVG 660
DB 601 PSGVPRDLSYMPIMKFPDEBEGACOPCINCHSCVDLDKGPAGORASPLTSYSAVVG 660
QY 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
DB 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILDEAVVMAGVSP 780
DB 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILDEAVVMAGVSP 780
QY 781 YVSRLLGICLTSTVOLVQMLPBYGCLLDHVRNRRGLSQDILLNMQOIAKMSYLEDVR 840
DB 781 YVSRLLGICLTSTVOLVQMLPBYGCLLDHVRNRRGLSQDILLNMQOIAKMSYLEDVR 840
QY 841 LVHRLAARNVLYKSPNHVKITDFGLARLDDIDETEHADGSKVPIKMALESILRRRT 900
DB 841 LVHRLAARNVLYKSPNHVKITDFGLARLDDIDETEHADGSKVPIKMALESILRRRT 900
QY 901 HOSDWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960
DB 901 HOSDWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVKKCM 960
QY 961 IDSECRPRRELVSFESPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLIEDDMGDLYNA 1020
DB 961 IDSECRPRRELVSFESPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLIEDDMGDLYNA 1020
QY 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 BEYLVPQOGFFCPCDDAPGAGVWVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDDPTVPLPSETDGYVAPLTCSPQPEXV	1140
Db	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDDPTVPLPSETDGYVAPLTCSPQPEXV	1140
Qy	1141	NQPVRRPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVYVDVFAFGAVENPEYLTPO	1200
Db	1141	NQPVRRPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVYVDVFAFGAVENPEYLTPO	1200
Qy	1201	GGAAPQPHPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQPHPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:40:39
Job time : 43.9774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds

(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-210-224-12

Perfect score: 6789

Sequence: 1 MELAALCRWGLLLALPPCA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: piri.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6698	98.7	1255	1 A24571	protein-tyrosine k
2	5880	86.6	1260	1 TVRTNU	protein-tyrosine k
3	5876.5	86.6	1254	2 I48161	p-185 precursor -
4	3076	45.3	1210	1 GORUR	epidermal growth f
5	3052	45.0	1210	2 A53183	epidermal growth f
6	3031.5	44.7	1223	1 TVCHLV	epidermal growth f
7	2909.5	42.9	1308	2 A47253	epidermal growth f
8	2609	38.4	1166	1 S06142	epidermal growth f
9	2345.5	34.5	1342	2 A36223	protein-tyrosine k
10	2260.5	33.3	1339	2 JG4387	kinase-related tra
11	1766.5	26.0	698	1 TVFVLV	epidermal growth f
12	1703	25.1	604	1 TVYUHL	protein-tyrosine k
13	1647	24.3	544	1 S35745	protein-tyrosine k
14	1640	24.2	545	2 S00727	protein-tyrosine k
15	1623	23.9	540	2 B44776	kinase-related tra
16	1621	23.9	540	1 TVFVVB	protein-tyrosine k
17	1569.5	23.1	1330	1 GQFBE	protein-tyrosine k
18	1444	21.3	644	2 A36325	epidermal growth f
19	1255	18.5	1323	2 E88257	epidermal growth f
20	1255	18.5	1374	2 S70712	protein let-23 [im
21	1175	17.3	1369	2 S70713	protein-tyrosine k
22	1155	17.0	1717	1 A45558	epidermal growth f
23	1063	15.7	527	2 A42032	epidermal growth f
24	914.5	13.5	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	epidermal growth f
26	754.5	11.1	311	2 T43220	protein-tyrosine k
27	675	9.9	1363	1 INHUR	insulin-like growt
28	675	9.9	1382	1 INHUR	insulin receptor p
29	666	9.8	1383	2 A36080	insulin receptor p

30	665.5	9.8	1372	2 A34157	insulin receptor p
31	662	9.8	1607	2 T43212	insulin-like growt
32	658.5	9.7	1300	2 A36502	insulin receptor-r
33	647	9.5	1268	2 B36502	insulin receptor-r
34	642	9.5	1477	2 T18534	protein-tyrosine k
35	603	8.9	1367	1 IGHUR1	insulin-like growt
36	595	8.8	987	2 A54092	protein-tyrosine k
37	595	8.8	1371	2 A33837	insulin-like growt
38	593.5	8.7	2148	1 A56081	insulin receptor -
39	589.5	8.7	977	2 S49004	tyrosine kinase Mp
40	589	8.7	987	2 I48652	mouse developmenta
41	588	8.7	1114	1 S05582	protein-tyrosine k
42	588	8.7	1390	2 T30346	insulin receptor -
43	588	8.7	2101	2 S57245	insulin receptor (
44	586	8.6	1091	2 S33596	protein-tyrosine k
45	584.5	8.6	976	2 A36355	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N.Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein e

C.Species: Homo sapiens (man)

C.Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #ext change 11-Jun-1999

C.Accession: A24571; A25491; A44188; B44188; I59509; I57622

R.Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T

Nature 319, 230-234, 1986

A.Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth

A.Reference number: A24571; MUID:86118663; PMID:3003577

A.Accession: A24571

A.Molecule type: mRNA

A.Residues: 1-1255 <YAM>

A.Cross-references: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:931198

R.Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A.Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid

A.Reference number: A25491; MUID:86016729; PMID:2995967

A.Accession: A25491

A.Molecule type: DNA

A.Residues: 737-1031 <SEM>

A.Cross-references: GB:M11767; NID:9182163; PIDN:AAA35808.1; PID:9553282

R.Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg

Science 230, 1132-1139, 1985

A.Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro

A.Reference number: A44188; MUID:86070181; PMID:2999974

A.Accession: A44188

A.Molecule type: DNA

A.Residues: 740-910 <COU>

A.Cross-references: GB:M12036; NID:9183988; PIDN:AAA35978.1; PID:9183989

A.Accession: B44188

A.Molecule type: mRNA

A.Residues: 1-517, 'RRL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU>

A.Cross-references: GB:M11730; NID:9183986

R.King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A.Reference number: I59509; MUID:85272597; PMID:2992089

A.Accession: I59509

A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 832-909 <REX>

A.Cross-references: GB:I29395; NID:9459807; PIDN:AAA35809.1; PID:9459808

R.Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcription

A.Reference number: I57622; MUID:87286698; PMID:3033351

A.Accession: I57622

A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-191 <TAL>

A:Cross-references: GB:M6792; NID:9183983; PIDN:AAA58637.1; PID:9553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics:
 A:Gene: GDB:ERBB2; NCL; NEU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:122-125/Domain: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:222-653/Domain: extracellular #status predicted <EXT>
 F:770-304/Domain: EGF receptor extracellular domain repeat <BE1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <BE2>
 F:654-675/Domain: transmembrane #status predicted <TMM>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68/124,187,259,571,629/Binding site: carboxylate (Asn) (covalent) #status predict
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.7%; Score 6698; DB 1; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 4,6e-279;
 Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MELALCRWGLLLALPPGAATVCTGDMKRLPASPETHLDMRLHLYGCGQVYQNL 60
 DB 1 MELALCRWGLLLALPPGAATVCTGDMKRLPASPETHLDMRLHLYGCGQVYQNL 60
 QY 61 ELYTPNNAISLFLODIOEVQYVLIANHQVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
 DB 61 ELYTPNNAISLFLODIOEVQYVLIANHQVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQORNPOLCYOPIILMKDIFHNQOLA 180
 DB 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQORNPOLCYOPIILMKDIFHNQOLA 180
 QY 181 LTLIDTNSRACHPCSPMKCKGRCSSESDCOSLTRVCAGAGCARCKPQYIKANSK 240
 DB 181 LTLIDTNSRACHPCSPMKCKGRCSSESDCOSLTRVCAGAGCARCKPQYIKANSK 240
 QY 241 FIGITELKSDCLACLFHNSGICELHCPALVTYNTDFESMPNBPGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNBPGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSPCARVCYIGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSPCARVCYIGMEHLREVRVTSAN 360
 QY 361 IOEPAGCKKIFGSLAFLESFDDGDPASNTAPLOPELOVFEYTEITGYLYISAMPDLSL 420
 DB 361 IOEPAGCKKIFGSLAFLESFDDGDPASNTAPLOPELOVFEYTEITGYLYISAMPDLSL 420
 QY 421 DLSVQNIQVIRGLILNAGVSLTIOGIGISWLGIRSLRELSGIALIHNTHLCPVHTV 480
 DB 421 DLSVQNIQVIRGLILNAGVSLTIOGIGISWLGIRSLRELSGIALIHNTHLCPVHTV 480
 QY 481 PMDQFRPHQALHTTANRPEDECVGEGLAQOLCARHGWCPGPTQCVNCSQFIRGDEC 540
 DB 481 PMDQFRPHQALHTTANRPEDECVGEGLAQOLCARHGWCPGPTQCVNCSQFIRGDEC 540
 QY 541 VEECRVILQGLPREYVNAHCLPCHPECOPNGSVTCFGEADQCVAAHYKDPFCVARC 600
 DB 541 VEECRVILQGLPREYVNAHCLPCHPECOPNGSVTCFGEADQCVAAHYKDPFCVARC 600
 QY 601 PSQVPRDLSYMPIMKFPDEEGACOPCPINCHSCVLDLDDKGPAPQASPLTISIYAVVG 660
 DB 601 PSQVPRDLSYMPIMKFPDEEGACOPCPINCHSCVLDLDDKGPAPQASPLTISIYAVVG 660

DB 601 PSQVPRDLSYMPIMKFPDEEGACOPCPINCHSCVLDLDDKGPAPQASPLTISIYAVVG 660
 QY 661 ILVVVLGVFGILILKRRQOKIRKTYMRLLOETELVEPLTPSGAMPNQAORILKETEL 720
 DB 661 ILVVVLGVFGILILKRRQOKIRKTYMRLLOETELVEPLTPSGAMPNQAORILKETEL 720
 QY 721 RKVYLGSAGFCTVYKGIIPDGENVKIPVAIKVLRNENSPYANKELIDEAVMAGVGSF 780
 DB 721 RKVYLGSAGFCTVYKGIIPDGENVKIPVAIKVLRNENSPYANKELIDEAVMAGVGSF 780
 QY 781 YVSRLLGICLTSTVOLVQLMPYGLLDHVRNRRGLSQDILNMCQIAKMSYLEYDR 840
 DB 781 YVSRLLGICLTSTVOLVQLMPYGLLDHVRNRRGLSQDILNMCQIAKMSYLEYDR 840
 QY 841 LVHRRLAARNVYKSPNHNKLTDFGLARLIDIEHYHADGKVPKMMALSLILRRFT 900
 DB 841 LVHRRLAARNVYKSPNHNKLTDFGLARLIDIEHYHADGKVPKMMALSLILRRFT 900
 QY 901 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMIWCKMM 960
 DB 901 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMIWCKMM 960
 QY 961 IDSECRPRRELVSFSSRNARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 DB 961 IDSECRPRRELVSFSSRNARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 QY 1021 EBYLVPQOGFCPPDPAAGGVNHRHSSSTRSGGDLTLGLEPSEEPAPSPPLAPSG 1080
 DB 1021 EBYLVPQOGFCPPDPAAGGVNHRHSSSTRSGGDLTLGLEPSEEPAPSPPLAPSG 1080
 QY 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOKRSSEDPVPLPSETGCVAPLCSQPEYV 1140
 DB 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOKRSSEDPVPLPSETGCVAPLCSQPEYV 1140
 QY 1141 NQPDVRPOPSPREBGPPLAARPAATLREPKTLSPGKGVADVAFGAVENPEYLTPQ 1200
 DB 1141 NQPDVRPOPSPREBGPPLAARPAATLREPKTLSPGKGVADVAFGAVENPEYLTPQ 1200
 QY 1201 GGAAPQPHPPPAFNDLNYWDDPBERGAPSTFKGPTPAENPEYLGLDVPV 1255
 DB 1201 GGAAPQPHPPPAFNDLNYWDDPBERGAPSTFKGPTPAENPEYLGLDVPV 1255

RESULT 2

TVRTNI

protein-tyrosine kinase (BC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence _revision 31-Dec-1988 #text_change 11-Jun-1999

C:Accession: A24562; A61204

R:Barzmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 <BAR>

A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746

R:Maatli, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen

Carcinogenesis 12, 1975-1978, 1991

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A:Molecule type: prelmimary

A:Status: prelmimary

A:Molecule type: DNA

A:Residues: 637-663, 'V', 665-702 <MAS>

A>Note: authors translated the codon GCA for residue 25 as Val

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:71,191,263,535,576,634/Binding site: carbonyldiester (Aaa) (covalent) #status predicted
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: Lys #status predicted
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.6%; Score 5880; DB 1; Length 1260;
 Best Local Similarity 86.7%; Pred. No. 4,2e-244;
 Matches 1090; Conservative 50; Mismatches 115; Indels 2; Gaps 2;

```

QY 1 MELAALCRWGLLALLPPGASTOVCTGDMKRLPASPEHLLDMLRHLYGCGVVOGNTL 60
DB 4 MELAAWCRWGLLALLPPGAGTGVCTGDMKRLPASPEHLLDMLRHLYGCGVVOGNTL 63
QY 61 ELTYLPTNASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDYALAVLNDG 120
DB 64 ELTYVPANASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDYALAVLNDR 123
QY 121 DPLNNTTPTVT-GASPGGLRELQSLTEILKGVLIQRPOLCYODTILMKDIFHKNNQL 179
DB 124 DPONVAASTPGRTPEGLRELQSLTEILKGVLIQRPOLCYODVLMKDVFRKNNQL 183
QY 180 ALTLIDNRSRACHPCSPMCKGSRCKGSSSDCSLTRTVCAGGACRCKPLPOYIKANS 239
DB 184 APVDIDNRSRACHPCSPMCKGSRCKGSSSDCSLTRTVCAGGACRCKPLPTDCCHEQ 243
QY 240 KFIITELKHSDDCLACHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTAC 299
DB 244 CAAGCTPRKSDCLACHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTTC 303
QY 300 PYNLTSDVGSCTLVCPPLNQEVTAAEDGTORCEKSKRCARVCYGLGMEHLREVAVTSA 359
DB 304 PYNLTSTEVGSCTLVCPPLNQEVTAAEDGTORCEKSKRCARVCYGLGMEHLRGARATSD 363
QY 360 NIGEFACKKTIFGSLALPLPESFDGDPASNTAPLQPELQVETLEETGYLYISAMPDSL 419
DB 364 NVEEFDECKKIFGSLALPLPESFDGDPASNTAPLQPELQVETLEETGYLYISAMPDSL 423
QY 420 PDLISFQNLQVIRIRLIHNGAYSLTLQGLISMTGLSLRLSGSLLIHNNHTLCFVHT 479
DB 424 RDLISFQNLQVIRIRLIHNGAYSLTLQGLISMTGLSLRLSGSLLIHNNHTLCFVHT 483
QY 480 VPMQDLFRNPHQALLHTANPEDE-CYEGELACHQLCARGHCMWGPRTQVCNCSQPLRQ 538
DB 484 VPMQDLFRNPHQALLHTANPEDE-CYEGELACHQLCARGHCMWGPRTQVCNCSQPLRQ 543
QY 539 ECVBECEVLOGLPREYVNAHRCLECHPECCQNGSVTCFGEALQCVACAHYKPPPCVA 598
DB 544 ECVBECEVLOGLPREYVNAHRCLECHPECCQNGSVTCFGEALQCVACAHYKPPPCVA 603
QY 599 RCEGVPRLSYMPIMKFPDEEGACOPCINCTSHCVDLDDKGPAPORASPLISYAV 658
DB 604 RCEGVPRLSYMPIMKFPDEEGACOPCINCTSHCVDLDDKGPAPORASPLISYAV 663
QY 659 VGLILVVLGVGILIKRROQKIRKXTMRLLQETELVPLTPSGAMPNOQMRILKET 718
DB 664 EGVLLFLLVVLGVGILIKRROQKIRKXTMRLLQETELVPLTPSGAMPNOQMRILKET 723
QY 719 ELAKVVLGSGAGCTYVKGIIWIPGENVKI PVALKULRENTSPKANKEIIDEAVMAGV 778
DB 724 ELAKVVLGSGAGCTYVKGIIWIPGENVKI PVALKULRENTSPKANKEIIDEAVMAGV 783
QY 779 SPYVSRLLGICLTSTVOLVQLMPYGLLDHVENRGLSODLLNMCQIAGKMSYLED 838
DB 784 SPYVSRLLGICLTSTVOLVQLMPYGLLDHVENRGLSODLLNMCQIAGKMSYLED 843
QY 839 VRLVHRDLAARNLVLSPNHVKITDQGLARLLDIDETEHADGGKVPDKMALESILRR 898
DB 844 VRLVHRDLAARNLVLSPNHVKITDQGLARLLDIDETEHADGGKVPDKMALESILRR 903
QY 899 FTHQSDVMSYGVTVWMLMTFGAKRPYDGI PAREILPDLLEKGERLPQPICTIDVYIMVNC 958

```

```

DB 904 FTHQSDVMSYGVTVWMLMTFGAKRPYDGI PAREILPDLLEKGERLPQPICTIDVYIMVNC 963
QY 959 WMIDSECRPRRELVSERMAADPQRFVUQNEPLGASPLDSFFYSLSLEDDMDGLV 1018
DB 964 WMIDSECRPRRELVSERMAADPQRFVUQNEPLGASPLDSFFYSLSLEDDMDGLV 1023
QY 1019 DAEELYVQOGGFFCDDPAPAGAGVHNRSSSTFGSGDGLTLGLEPSEEARSPPLAS 1078
DB 1024 DAEELYVQOGGFFCDDPAPAGAGVHNRSSSTFGSGDGLTLGLEPSEEARSPPLAS 1083
QY 1079 EGAGSDVFDGDLGMAKGLQSLPETHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQE 1138
DB 1084 EGAGSDVFDGDLGMAKGLQSLPETHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQE 1143
QY 1139 VYNOEDVPRPPSPREGVLPAPARAGATLEBAKTLSPEKGVVMDVFAFGAVENPEYLT 1198
DB 1144 VYNOEDVPRPPSPREGVLPAPARAGATLEBAKTLSPEKGVVMDVFAFGAVENPEYLT 1203
QY 1199 POGGAAPQPPPPAFSPAFDNLVYWDODPPRSGAPSTFFKGPPTAENPEYLGLDVVP 1255
DB 1204 PREGTASPRHSPAFSPAFDNLVYWDODPPRSGAPSTFFKGPPTAENPEYLGLDVVP 1260

```

RESULT 3

148161
 P-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C:Accession: 148161
 R:Nakamura, T.; Uchijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
 Gene 140, 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: 148161; MUID:94193007; PMID:7908275
 A:Accession: 148161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:g493236; PID:BAA03801.1; PID:g747595
 A:Genetic:
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.6%; Score 5876.5; DB 2; Length 1254;
 Best Local Similarity 86.5%; Pred. No. 5.8e-244;
 Matches 1086; Conservative 58; Mismatches 110; Indels 1; Gaps 1;

```

QY 1 MELAALCRWGLLALLPPGASTOVCTGDMKRLPASPEHLLDMLRHLYGCGVVOGNTL 60
DB 1 MELAAWCRWGLLALLPPGAGTGVCTGDMKRLPASPEHLLDMLRHLYGCGVVOGNTL 63
QY 61 ELTYLPTNASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDYALAVLNDG 120
DB 64 ELTYVPANASTLFLQDIOEVQGYVLIANOVROYPLQRLIVRGTOLEFEDYALAVLNDR 123
QY 121 DPLNNTTPTVTGASPGGLRELQSLTEILKGVLIQRPOLCYODTILMKDIFHKNNQLA 180
DB 124 DPLDNVTATGRTPEGLRELQSLTEILKGVLIQRPOLCYODVLMKDVFRKNNQLA 180
QY 181 LTLIDNRSRACHPCSPMCKGSRCKGSSSDCSLTRTVCAGGACRCKPLPOYIKANSK 240
DB 184 LTLIDNRSRACHPCSPMCKGSRCKGSSSDCSLTRTVCAGGACRCKPLPOYIKANSK 240
QY 241 FGIITELKHSDDCLACHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTAC 300
DB 244 AAGCTPRKSDCLACHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCYTTC 300
QY 301 PYNLTSDVGSCTLVCPPLNQEVTAAEDGTORCEKSKRCARVCYGLGMEHLREVAVTSA 360
DB 304 PYNLTSTEVGSCTLVCPPLNQEVTAAEDGTORCEKSKRCARVCYGLGMEHLRGARATSA 360

```

Qy	361	IOEFAGCKKIFGSLAFIPESFDGPBASTAIOEOLOVEFTLEIICYLISAMPOLP	420
Dp	361	IOEPAQCKKI FGSIAFLPESFPDGNSSIAPLTPOLOVFEFTLEIICYLISAMPOLH	420
Qy	421	DLASFONLQVIRGRILHNHGAYSLTLPQIGISIMLGRSLRELSGSLALHNHTLCPVYTV	480
Dp	421	DLASFONLRVIRGRVILHDGAVSLALQOGIGIMLGRSLRELSGSLVILHRNTHLCFYHTV	480
Qy	481	PMDOLFRRPHOALLHTANRPDECEVGBELACHOLCARHCHMGPRPTQCVNCSOTLRGOC	540
Dp	481	PMDOLFRRPHOALLHSGNPSSEEGKLQFACYPICAHCHMGPGPTQCVNCSHFLRGQEC	540
Qy	541	VEECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFPEADOCVACHYDDPFCVYAR	600
Dp	541	VKECRVWKGJLPREYVNGKCLPCHPRECOPQNSTETCTGSEADQCTACHYDPSFCVYAR	600
Qy	601	PSGVKPDLSYMPIMKFPDEBEGACOPCIPINCHSCYVDLDDKCCPAEORASPLTISVAVG	660
Dp	601	PSGVKPDLSYMPIMKYDPDEBGMCPCPICINCHSCYVDLDERCCPAEORASPATSIATVVG	660
Qy	661	ILLVVLGVNVGIIILKRQOKIRKTYMRLLQETLVEPLTPSGAMPQAOMRILKETEL	720
Dp	661	ILLELVIGVNVGIIILKRROKIRKTYMRLLQETLVEPLTPSGAMPQAOMRILKETEL	720
Qy	721	RKVYVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLENTSPKANKEIIDEAYVYAGVGP	780
Dp	721	RKVYVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLENTSPKANKEIIDEAYVYAGLGSP	780
Qy	781	YVSRLLIGCLTSTVOLVTLMPYCGLLDHNENGRGSGOPLMMCMQOIAAGMSYLEBYR	840
Dp	781	YVSRLLIGCLTSTVOLVTLMPYCGLLDHNVEHHRRLSGOPLMMCVQOIAAGMSYLEBYR	840
Qy	841	LVRHDLAARNLVYSPNHVKITTFGLALBIDETEVYADGCAKVIKMMALESI LRREFT	900
Dp	841	LVRHDLAARNLVYSPNHVKITTFGLARLIDETEVYADGCAKVIKMMALESI LRREFT	900
Qy	901	HOSDVMSGVTVVWELMTFGAKPYDCIPARELPDLLEKEERLPORPICITIDVYMMYKCM	960
Dp	901	HOSDVMSGVTVVWELMTFGAKPYDCIPARELPDLLEKEERLPORPICITIDVYMMYKCM	960
Qy	961	IDSECRPFRELVSEFSRMARDPQRFVVIQONEDLGPAPLDSTFASRLSEDDMDGDLVDA	1020
Dp	961	IDSECRPFRELVSEFSRMARDPQRFVVIQONEDLGSPSLDSTFASRLSEDDMDGDLVDA	1020
Qy	1021	EYVLVPOQGFPCPDPAFAGAGMWHHRHSSSTRSGGDLTLCLEBSEEARSPLABSEG	1080
Dp	1021	EYVLVPOQGFPCPDPAFAGAGSTAHRRHSSSTRSGGGLTLCGMEPSGEEPRSPLABSEG	1080
Qy	1081	AGSDVFODDLGMGAAGKLOSLPTHNDPSLOYSEPTVPLPSETDGYVAYPLTCSPOPEYV	1140
Dp	1081	AGSDVFEELGMGATKGPQISISPRDLSLOYSEPTPLPSETDGYVAYPLACSBQPEYV	1140
Qy	1141	NOPDVRPOPSPREGPLPAARPAGATLERAKTILSPKNGVYKDYFAFGAVENBEYLTPO	1200
Dp	1141	NOPEVRPOPLTPREGPLPVVRPAGATLERPTILSGKNGVYKDYVTFGAVENBEYLVPR	1200
Qy	1201	GGASPOHPPAPAFSAPFNDLYYMODDPERGAPSTFPGCTPAENPEYLGIDVPU	1254
Dp	1201	GGASPOHPPALCPAFNDLYYMODDPERGAPSTFPGCTPAENPEYLGIDVPU	1254

RESULT 4
 GENE
 epidermal growth factor receptor precursor - human
 N:Contains: protein-tyrosine kinase (BC 2.7.1.112) erbB
 C:Species: Homo sapiens (man)
 C>Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
 C:Accession: A00611, R25772; S30024; A38622; A00642; A43615; A23062; A05281; A60143; A33
 R:Rullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
 19, P.H.
 Nature 309, 418-425, 1984
 Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression ob

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <UUL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA5240.1; PID:g757924
A:Note: The authors translated the codon AAG for residue 540 as Asn
R:Smith, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene
A:Reference number: A25772; MUID:85270438; PMID:2591899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <IGH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA5370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification of the promoter
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 17461-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termination in the regulation of the human epidermal growth factor receptor gene
A:Reference number: A38672; MUID:91107677; PMID:1588448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA6171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Waterfield, M.D.
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330553
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCMRA', 150-187, 'KSVQNAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-327, 'T', 329-353, 'T', 355-387, 'T', 389-411, 'R', 413-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF receptors
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Scolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.M.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification and characterization
A:Reference number: A43615; MUID:84196372; PMID:6326251
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
A:Reference number: A60143; MUID:85182650; PMID:2985560
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUB>
R:McCoy, B.; Mosis, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
A:Reference number: A58023; MUID:84191554; PMID:6325948
A:Accession: A58023
A:Contents: annotation; receptor activity

A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A:Title: Functional independence of the epidermal growth factor receptor from a domain
 A:Reference number: A3331; MUID:9000223; PMID:2790960
 A:Content: annotation: internalization signal
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
 C:Gene: EGFR
 A:Gene: GDB:120610; OMIM:131550
 A:Cross-references: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP, autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-110/Domain: EGF receptor #status predicted <MAT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F:646-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
 F:745/Active site: Lys #status experimental

Query Match 45.3%; Score 3076; DB 1; Length 1210;
 Best Local Similarity 49.0%; Pred. No. 2,5e-124;
 Matches 620; Conservative 177; Mismatches 362; Indels 106; Gaps 21;

11 LLLALLPPGAA--STOVCTGTMKRLPASPTHLLMLRLLYOGCQVVGNTLPTN 68
 14 LLLALLPPGAA--STOVCTGTMKRLPASPTHLLMLRLLYOGCQVVGNTLPTN 73
 69 ASISFLODIOEVGVYLAHNOYRQVLRQRLRYGQLFEDNYALAVDNGPLNATTP 128
 74 YDSFRTKTOEVGVYLAHNOYRQVLRQRLRYGQLFEDNYALAVDNGPLNATTP 126
 129 VTGASPGRLRELQRLSTELLKGVLLIQRNPOLCYODTILMKDIFHKNOALATLIDTNR 188
 127 ---ANKTGLKELPMRLNDELHGAVRSNNPALCNESIQMRDVSDFSNMSMDQNH 183
 189 SRACHPCSPMCKSKRCGSSSEDCQSLTRTVCAAGCA-RCKGRLPOYIKANSKFTIGTEL 247
 184 LSGCQKCDSPGSCPGAGBENCQKLTIKIICAOQCSGRCGRKSPDCCNQCAGCTGP 243
 248 KHSDDLACLFHNSGICELCPALVYNTDFFESMRPRBGRYTFGASCYATCPYNYISTD 307
 244 RESDCLVCRKFRDEATCTKDTCPPLMLNPTTYQMDVNEBKYSFGATCVKCCPRMYVYTD 303
 308 VGSCTLVCPRLHNOEYVAEDTORCEKSKPCARVCYGLGMEHLREYVAVTSANIOEPAGC 367
 304 HGSQVACAGADSYEM-EBDGVKCKKCEGCRKVCNGIIGERKDSISIAATNIKFKNC 362
 368 KIFGSLAFLPESFDDDPASNTAPLOPQLQVEFTLEITGYLYISAMPDLDLSVFN 427
 363 TSIISGLHLPLVAFRSDSFHTPRLDQELDIKTVETIGFLIQAEMPRNRTDLHAFEN 422
 428 LQVIRGRILHNGAYSLTLOGISWLGSLRLSGALILHNTHLCPHTYPMQLFR 487
 423 LELIRKRTQKHGFSIAVSLNLTSLGLSLKLSIGDVLISGNKMLCYANTINMKLFCG 482
 488 NPHQALHTANRDEDECVGEGLAGHOLCARGHCMPGPTQVACSOFLRQECVCEGRYL 547
 483 TSGQKTKIISNGENSCAKTQVCHALCSPEGCMGEBPRVCYCRVNSRRECVADCKLL 542
 548 QGLPREYVNAKRLPHRPGOPNGSVTCGPRADOCVACAKDPRFVACRPSGVKPD 607
 543 EGPREFVENSEICQHPCLPQANNITCGRPRDNCIOCAHYIDGRHCVKTCRPAVMG 602
 608 LSYMPKPRDEGACQPCPINTGSHCVLDLDDGCPAEGRASPLTSTIVAVVG---ILLV 664

Db 603 NNTL-VWKYADAGVCHLCHENCTYCTGPGLEGCPNTGPKIP--STATGVGALLLLV 659
 Oy 665 VLVGVGVGILIKRQOQIKRKYTKMRLLQETLVPPLTPSGAMPQAOQRILKETELRVK 724
 Db 660 VALDQ---LPMRRRIYRKKTLLRLQERLVEPLTPSGAPNALRLIKETEFKKIK 716
 Oy 725 VLSGAGFYVKGIMIPDGENVKIPVAIKVLENTSPKANKEILDEAVYMGVSPVYSR 784
 Db 717 VLSGAGFYVKGIMIPDGENVKIPVAIKVLENTSPKANKEILDEAVYMGVSPVYSR 776
 Oy 785 LLSGLTSTYQVLVQLMPYGLLDHYRBNKRLCSODLLMKCMQIAKMSYLEDRVLVR 844
 Db 777 LLSGLTSTYQVLVQLMPYGLLDHYRBNKRLCSODLLMKCMQIAKMSYLEDRVLVR 836
 Oy 845 DLAAKNTLVSPNPKITDGLARLLIDEPYVADGKVYIKMALESILRRPFTQSD 904
 Db 837 DLAAKNTLVSPNPKITDGLARLLIDEPYVADGKVYIKMALESILRRPFTQSD 896
 Oy 905 VMSYGVTVLWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIKCMIDSE 964
 Db 897 VMSYGVTVLWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIKCMIDSE 956
 Oy 965 CRPRFRLVSEFSPMAADPQRFVYIQ-NEDLGRASPILDSTFYRLSLDDDGDLVDAEY 1023
 Db 957 SRPFRLIIEFSGKARDPQRYLVIQGDERMHLPSPTDSNFRALMDEEDMDVDADAEY 1016
 Oy 1024 LVPOQGFPCDPAAGAGMHHRRSSSTSGGDDLTLGLEPSEEARPSPLASEGAS 1083
 Db 1017 LVPOQGFPCDPAAGAGMHHRRSSSTSGGDDLTLGLEPSEEARPSPLASEGAS 1042
 Oy 1084 DVPDGLGMAKGLASLPDHPDSPLQRYSEDPTVPPLSET--DGYVAPLTCSPQPEVYN 1141
 Db 1043 N--NSTVACIDRNGLOSCPIKEDSFLQRYSDPTGALTBDSIDDTPL-----PVPEYIN 1094
 Oy 1142 QPDVRFQPSRRCBGLPAPRAGATLERAKTISFGKNVYKVYVAFGAVENPEYL-TPQ 1200
 Db 1095 Q-SVPRKRPAGSVNPPVYHNPQPLNP-----APSRDPHYQD--PHSTAVGNPEYLVQ 1143
 Oy 1201 GGAAPQHPAPSPAFNDLYWMDQ-----DP-----PERAPSTFGCTPAE 1244
 Db 1144 -----PTCVNSTFDSPAHMAQKSHQISLDNPYQODFFKAKPNGIIFKGS-TAE 1193
 Oy 1245 NPEYL 1249
 Db 1194 NAEYL 1198

RESULT 5

A53183

Epidermal growth factor receptor precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence: revision 06-Jan-1995 #text: change 18-Jun-1999

C:Accession: A53183; A43818; S24942; A28941; S45325; I49643

R:Luteteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.

Genes Dev. 8, 399-413, 1994

A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor

A:Reference number: A53183; MUID:94170986; PMID:8125255

A:Accession: A53183

A:Molecule type: mRNA

A:Residues: 1-1210 <LNR>

A:Cross-references: GB:U03425

R:Aviv, A.; Iax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.

Oncogene 6, 673-676, 1991

A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding

A:Reference number: A43818; MUID:91232866; PMID:2030916

A:Accession: A43818

A:Molecule type: mRNA

A:Residues: 1-714 <AVI>

A:Cross-references: GB:X59698

R:Risinger, D.P.; Serrero, G.

submitted to the EMBL Data Library, June 1992

A:Reference number: S24942

A:Accession: S24942

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:212608
R:Heiserfmann, G.J.; G. J.; G. N.
J. Biol. Chem. 263, 1152-1158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hubb, M. L.; Dunn, A. R.; Alexander, W. S.
Submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: 545325
A:Accession: 545325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CA55587.1; PID:9488831
R:Paria, B. C.; Das, S. K.; Andrews, G. K.; Dey, S. K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: 149643; MUID:93126380; PMID:7678348
A:Accession: 149643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:L06864; NID:9193001; PIDN:AA553029.1; PID:9567201
C:Genetic8:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TMM>
F:712-977/Domain: protein kinase ATP-binding motif
F:720-728/Region: protein kinase ATP-binding motif
F:689,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1157/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.0%; Score 3052; DB 2; Length 1210;
Best Local Similarity 49.0%; Pred. No. 2, 7e-123;
Matches 623; Conservative 169; Mismatches 370; Indels 110; Gaps 23;

11 LLLALLPQAA--STOVCTGDMKLRPASPTHLDMRLHYOGQVVOGNETLPTN 68
14 LLLALCAAGALEEKVKVCGTSMRLTOLGTFEDHFLSLQRMYNCEVVLGNLEITYQRN 73
69 ASLSFLDIOEVQGVVLAHNOVROVPLORLRIVRGTOLEFEDNYALAVLNDPLNNTTP 128
74 YDLSLKTIOEVAGVVLALNTVERIPLBNLOIRGNALYENTYVALATLSN----- 124
129 VTGASPGGLRELQSLTEILKGVLIQRNPOLCYODTILMKDI----FKKNQALTLI 184
125 -YGTNRGTGLRELPMNLDELILGAVRFSNNPILCMMDTIQWRDIQVNFMSMSMDL--- 180
185 DTNSRACHPPSGPMKSGRCWGESESDQSLTRTYCAGCA-RCKGPLYQYIKANSKRTIG 243
181 -QSHPSSCPCKDPPGNSCWMGGGEGENCOKLTKIICAOQCPSHRCGRSPSDCHNOCAAG 239
244 ITELKHSPLACLNHNSGICELHCPALVTYNTDIFESMPNREGRYTGASCVTACPNY 303
240 CTGPRESCLVCKQKODEATCTKOTCPPLMLNPTTYQNDVNBEGRYSFATCVKCPNRY 299
304 LSTDVGSCTLVCPNLNOEVTADGTQRCESKSPCARVCYGLGHEHLREVRATSIQOE 363
300 VVTDHGS CVRACGPRYEV-EEGDIRKCKCKCGPRKXVNGSIGIEFDTLSINNTIKH 358
364 FAGCKKTIGSLAFLEPSDGDPRASNTAPLOPQOLQVFTLEITTYGLYISAMPDLPLIS 423
359 FKYCTAISGDHLTLVAFKGDSTFRTPLDRELLEILKTVKEITGFLILQAMPDMWTDLH 418

424 VPOQLVTRGLTHNGAVSLTLOGISIMLGRSLRELGLALITHNTHLCFVHTVPKD 483
419 APENLLEIRGRKQKQGFSLAVGLNITSLGRSLKEISDGVITISGNNLCLVANTINRK 478
484 QLEFRPHOALLHTANRPEDECVEGLAQHOLCARGHCWCPGPGTQCNCQSOLRGOECVE 543
479 KLFGTPNQKTKIMNRAEKDCAVNVHVCNPLCSSGCMQPERDVCSCNVSRGRGECVEK 538
544 CRVLQGLPREYVNAHCLPCHPECPONGSVTCFPEADQVACAHYKDPFCVACRPSG 603
539 CNILEGEPRFEVNESECIOCHPECLPQAMNITCTGCRGDNICQAHYIDGPHCVTKCPAG 598
604 VKPDLSTWPIKFPREBEACQCPPTNCHSCVYDLDDKGPAGQASPLTSYSAVVGLL 663
599 IMGENNTL-VWKYADANVCHLCHANCYTGCGPGLQCEVWPSGPKIPSLATGIVGLL 657
664 VVVLQGVFGL-LIKRQOKIRKYTRRLRLQETELVEPLTPSGAMPNOAMRIKETELRK 722
658 FIVV-VALGIGLFMRRIHVKRTLRLLQRELEVEPLTPSGEARNQALRLIKETERFK 716
723 VKVLGSAFGVYKGIWIPDGENVKI PVAIKVLRNTPSKANKELIDEAYVAVAGSPVY 782
717 IKVLGSAFGVYKGLWIPDGENVKI PVAIKELREATSPKANKELIDEAYVAVAVONPVH 776
783 SRLIGICTSTVOLVTOQLMPYCCLDHRENRGRGLGSDLLNMCQOIAKMSYLEVRLV 842
777 CRLIGICTSTVOLVTOQLMPYCCLDHRENRGRGLGSDLLNMCQOIAKMSYLEVRLV 836
843 HRDLAARNVLYKSPHNVKITDPGLARLIDIDETEHADGKVPIMMALESILRRPFTHQ 902
837 HRDLAARNVLYKTPHNVKITDPGLARLIDIDETEHADGKVPIMMALESILRRPFTHQ 896
903 SDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDVYMIKCMWID 962
897 SDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDVYMIKCMWID 956
963 SECRPRFELVSEFRMADPQRFVYIQ-NELGASPLDSFTFYSLLEDMDGDLVDE 1021
957 ADSRKRFELVSEFRMADPQRFVYIQ-NELGASPLDSFTFYSLLEDMDGDLVDE 1016
1022 EYLVPOQGFPCDPPAGAGMWHHRSSSTRSGGDLTLGLPESEEARSPPLASEGA 1081
1017 EYLVPOQGFPCDPPAGAGMWHHRSSSTRSGGDLTLGLPESEEARSPPLASEGA 1042
1082 GSDVFDGLGKCAKAGLQSLPTHDPSPLQRYSEDPVLPSET--DGVAAPLTCSPQPEY 1139
1043 TSN-----NSTVACINRNGSCRYKEDAFLORYSSDPTGAVTBDNIDDAFL-----PVPEY 1092
1140 VNQPPVRRQPPSRGCPPLPAARPAATLEAKTILSPGKNGVYKDVAFAGAVENPEYL-T 1198
1093 VNQ-SVPRKPAQSVQNPVYHNPPLHP-----APGDLHAYQ--PHSNAGNBEYLMT 1141
1199 POGGAAPRHPPPAPSPAFDNLVYWDQ-----DP-----PERGAPSTFGPT 1242
1142 AQ-----PTCSSGFSNPAIMIQGSHMSLNDPNYQODFFPKETKPGITRG-PT 1191
1243 AENPEYGLDVP 1254
1192 AENPEYGLDVP 1203

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C>Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Max, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Venn,
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in
A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 A:Cross-references: GB:M20386
 R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroidblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:8522822; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M10066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 specific protein kinase
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:81-307/Domain: EGF receptor extracellular domain repeat <EB1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <EB2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.7%; Score 3031.5; DB 1; Length 1223;
 Best Local Similarity 48.0%; Pred. No. 2e-122;
 Matches 622; Conservative 174; Mismatches 356; Indels 145; Gaps 25;

8 RWGLLLLPFGA-----STVCTGDMKRLPASETHLDMRLHYGCGVQVQGLE 61
 13 RGAALVLLGLVALLGALCSAVEEKVCGGTNNKLTQLGHVEDHTSLQRYMNCVLLSYLE 72
 62 LTVPTNASTLFLODIOEVGVVLIANQVRQVLPORLRIVRGTLQFEDNYALAVLDNGD 121
 73 ITVEHNRDLTFLKTIQVAGVLIAMNDVIFLENQIIRGNVLYNSRALLAVSYH 132
 122 PLNNTTPVSGPGLRELRSLTEILKGGVLIQRNPOLCYODITLWKDIFHKMQLAL 181
 133 -MNTQ-----GLRELPMKRLSEILNGVKISNNPKLCNMDVLENDIIDSRL-PL 182
 182 TLID-TNRSRACHSPKCKSRCKGSESDCCSLTRVCAAGCA-RKGFPLQYITKANS 239
 183 TVLDFASSLSCPCPKPCTEDHCWAGEONCOTLTKVICAQCSGRCRGVVPSCCHNQ 242
 240 KFIGITELKHSDDCLAFHNSGICELCPALVTYNTDTPFSPMPREGRYFGASCVTAC 299
 243 CAAGCTGPRESDCLAKRFRDATCKOTCPPLVLYNPTTYQMDVNPBESKIFGATCAEC 302
 300 PYNLTSDVSGCTLVCPRLNQEVTADGTORCEKSKPCARVYGLGMEHLREVAVYSA 359
 303 PNNVTVDHGSCVASCNDTYEV-EENGVKCKKCDGLCSKVCNGIGIGELKGLISNAT 361
 360 NIGFACKCKTFGSLALPESFDDPASNTPLPBOLOVETLEITGYLYISAMPDSL 419
 362 NISFCKCTKINGVSTLPAVFLDPAFTKPLDLPKCLDVRTYEISGFLLIQAMPDNA 421
 420 PDLVFNOLQVIRRIIHNQAYSLTLOGISWTLGSLRLRELSGLLIIHNTHLCFVHT 479
 422 TDLYAFENLEIRKTRKHOGYSLAVNVLKISGLSLKISIGDIAIMOKKYLCTADT 481
 480 VPMDOLEFRNPHQALLHTANREDECVBEGLAHQOLCARGHCGWGPPTQVNCOSQFLRQE 539
 482 MNRSLEFATOSQKTKIIONRNKNDCTADRHVCDBLCSDVGCWGPFPFCSCRFSSQKE 541
 540 CVEECRLQGLRPRYVNAHGLPCHPEQPNG---SVTCGRPADQCVACAHYKDPFC 596

DB 542 CVKQCNIIQGEPRERERDSKCLPCHSECLVONSTAVNTTCSGCPDHCCKCAHFLDGPNC 601
 597 VARGSGVAPDLSTVPIKFPDEGACOPCPNCTHSCVDLDDKCGPAGORASPLTSIYS 656
 602 VKACRAGVLGENDTL-VKRYADANAVALCQCHNCTRGCKGPELEGP--NSGKTPSIA 657
 657 AVV-GILLVVVLGVVFGLIKRQOKIRRYTRRLLOETELVEPLTPSGAMPNOAMRL 715
 658 GVGGGLCLLVVVGIGIYLRRR-HIYKRTLRLQRELVETLTPSEARNQHLRL 716
 716 KETELARKVULGSAFGVYVYGIKIPDEBNKIPVAIKVLRNTPSPKANKELIDEAYVMA 775
 717 KETEFKVKVLSGSAFGVYVYGLMIPBEKVKIPVAIKELREATSPKANKELIDEAYVMA 776
 776 GVGSPPVSRLLGICTSTVOALMPGCLLDHRENGRGLSGODLLMWCKQIAKMGVY 835
 777 SYDNPHVRLGLICTSTVOALMPGCLLDHRENGRGLSGODLLMWCKQIAKMGVY 836
 836 LEDVRLVHRDLAARNVLKVSBNVKITDPGLARLLDIDETEHADGKVPIMKMALESIL 895
 837 LEERLVHRDLAARNVLKVTQHVKITDPGLAKLGADEKEVHAEGKVPIMKMALESIL 896
 896 RRRFTGSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
 897 HRIYTHOSDVSYGVTWELMTFGSKPYDGIIPASEISSVLEKGERLPQPICTIDVYIM 956
 956 VKCMNIDSECPREPELVESESRMARDPORFVVO-NEDLGPASPLDSTFYSLLEDMD 1014
 957 VKCMNIDSECPREPELVESESRMARDPORFVVO-NEDLGPASPLDSTFYSLLEDMD 1016
 1015 GDLYAEVLYPOOGFFPCDPAPGACGAVNHRHSSSTRSGGDLTLGLPSEEEAPRSP 1074
 1017 EDYDADYLVPHQGF-----NSPST-----SRTP 1042
 1075 L-----APBEGASDVPFDGLAMGAKGLSLPTHDSPLORYSEDPVLPSEF--DGY 1127
 1043 LLSSLSATSNNSATNCTID-----RNGQGHPRVEDSFVORYSSDPGNFLEESIDGG 1094
 1128 VAPLTGSPQREYVNOVDVRPQPSRREGRLPAARPAAGTLERAKTLSCGKGVVADV-- 1185
 1095 L-----PAPRYVNO--LMPKPS-----TAVVNOQIYNNISLT 1125
 1186 -----AFGAVNPEVLTPOGGAAPQHPAPFAPFADNYVMDQ----- 1225
 1126 AISKLPMSRQNSHSTAVNDPEYL-----NTNOSPRLAKTVPESSPYMIQSGNHOIN 1177
 1226 -DPE-----KAPSTFKGTPTAENPEYLGIDVP 1254
 1178 LDNPDYOODPLPNETKPMGLLKVAENPEYLRVAAP 1214

DB 1226 -DPE-----KAPSTFKGTPTAENPEYLGIDVP 1254
 1178 LDNPDYOODPLPNETKPMGLLKVAENPEYLRVAAP 1214

RESULT 7
 A47253
 epidermal growth factor receptor, HER4 - human
 C:Species: Homo sapiens (man)
 C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
 A:Accession: A47253
 R:Plowman, G.D.; Couloucou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
 A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderm
 A:Reference number: A47253; MUID:93189574; PMID:8363326
 A:Accession: A47253
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1308 <ELO>
 A:Cross-references: GB:I07868; NID:9337359; PIDN:AA59446.1; PID:9337360
 A:Note: sequence extracted from NCBI backbone (NCBI:126842)
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor
 F:716-981/Domain: protein kinase homology <KIN>
 F:724-732/Region: protein kinase ATP-binding motif


```

Db      123 YQK-NPSSP--DVTQVGLKQULSNTLFTLSGCVKSHNPLLCVETINMMDIVDKTSNP 179
Qy      180 ALTLIDTNRSRACHPCSPMCKSGRCWSESSDCSLTRTVAGGC-ARCKGRLPOLYIKAN 238
      180 TMLNLIPIHAFERQCKCHGICVNGSCWAPRGHCQKFTLLCAEQCNRCRCKPRLDCNE 239
Qy      239 SKFGLITELKHSDDLCLLHFHNSGICELHCPALVTYNTDTESMPNEGRYTFGASCYTA 298
      240 HCAGCGTGPRAITDLACGRDFNDGCKTCKPTKIDIVSHQVNDPNIKYTFGAAVCYE 299
Qy      299 CPVYVLTSDVGSCTLVCPLNQEVTAEDGTORCEKSCPCARVCGIGMEHLREVRATYS 358
      300 CPSPVYVTE-GACVRSRCSAGMLEVD-ENKGRSCCKPCDGVCKVCDGIGISLSTIANVS 357
Qy      359 ANIOEFAQCKKIFGSLAFLPESPFDGPASNTAPLOEOLQVFETLEITGYLYISAMPDS 418
      358 TNIRSFNCTKINDIILNRSFEGDPHYKIGTMDPEHLMVLTYYKEITGYLVIMMWEEN 417
Qy      419 LPDLVSQNLQVIRGRILHNGAYS-LTLQGLIGISMLGRSIRELSGSLALIHNTHLCEV 477
      418 MTSLSVFQNLLEIRGRITFSRGSFVVQVNRHQLMGLIRSLKEVSAGVILIKNTLQLRYA 477
Qy      478 HTVPMDDLFRNPHOALLHTANRPDECEVGEGLACHQICARHCWGPRTQCVNCSQPLRG 537
      478 NTIMWRFLFRSESDOSIETDART-----ENQTCNNECESDGCW-PGPTWCVSCLIHYDRG 529
Qy      538 QECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCAACHYKOPFCV 597
      530 GRCVASCNLQGEFREAOVORCVQCHOECLVQDLSITCYVGRANCSKSHFDDGQCI 589
Qy      598 ARCPGSKPDLSTWPIKFPDEBACOPCPINCHSCVDLDDKCCPABQRPASPLTISVA 657
      590 PRCPHGILGDDTL-INKYADIKMGQCCPCHQNTQSGSPGLSGCRDP-IYSHSLAVGL 647
Qy      658 VVGLLVVVLGVVGLIKRROOKIRKYTWRLLOETVEIPELTPSGAMPQAOAKRIKE 717
      648 VSGLLIVIVIALLVLLRRRRRIK-RKRTIKCLQEBELVEPLPSQAPQOALRIKE 706
Qy      718 TELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKAKEILDEAYVAGV 777
      707 TERKQORVLGSGAFGVYKGLMNDGENINIRPVAIKVLRNTPSKNOEVLDEAYVAVS 766
Qy      778 GSYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENRGRIGSODLWMCQIAKMSYLE 837
      767 DHHVHVRLLGICLTSAVQVLTQMLPYGCLLDYVQHOERICGQWLNMVQIAKMSYLE 826
Qy      838 DVLVHRDLAARVNLVSPVHVKITDGLARLDDIDETEHADGSKPIPMMALESLTR 897
      827 ERHLVHRDLAARVNLKPNHVKITDGLSKLTADEKEYQADGKVPKIMMALESLTOW 886
Qy      898 RFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPIDLLEKGERLPOPICTIDVYMIWK 957
      887 TYHQSDVMSYGVTVWELMTFGSKPYDGIPIAKEIASLWENGERLPOPICTIEYMIILK 946
Qy      958 CMWIDECRRPRELYSEFSKMARDPORFVLIQNEIDIGPASPLDSTFYRSLDEDDMGDL 1017
      947 CMWIDSSRRPRELYSEFSQMARDPKRYLVIQ--NLPSLSDRLFRSLSSD--DV 1001
Qy      1018 VDAEETLVPOQGFCCPDPAAGAGWVHHRRSSSTSGGDLTLGLEPSEEARPSRLAP 1077
      1002 VDADELTLPRKRL-----NMQGS-----EPCIP 1024
Qy      1078 SEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPV-DLPSETDGYVAPLTCSPQ 1136
      1025 PTGH-----PVRENSITLRNLSDPQLMLEDDOH----- 1055
Qy      1137 PEYVNPDPVRPQ-----PSPRE-----GPLP-AARPAATLEBAKTILSPKNGVYKD 1183
      1056 -EYVNPQGSSTSRSLSDIYNPNEDLTDGWPVSLSSOEATNFSPEYINTQNSL--- 1111
Qy      1184 VFAFGAVENPEYLTPOGGAAPRHPRAPSPAFNDLYWDDQPRBERGAPPSFKGTPTA 1243
      1112 PLVSSGSMDDPY---QAG-----YQAAV-----LPQYTGALTNGMFLPA 1149

```

```

Qy      1244 ENPEYIG 1250
      1150 ENLEYIG 1156

RESULT 9
A36223
kinase-related transforming protein (erbB) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kusl, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal
A:Accession: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <RFA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-
A:Reference number: 159164; MUID:90311312; PMID:2164210
A:Accession: 159164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559,'G',561-957,'F',959-1063,'G',1065-1342 <RES>
A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match      34.5%; Score 2345.5; DB 2; Length 1342;
Best Local Similarity 40.0%; Pred. No. 4.6e-93;
Matches 525; Conservative 190; Mismatches 467; Indels 129; Gaps 32;

Qy      10 GLLALLPRGA--STQVCTGDMKRLRPSPEHLDMLHLYOGCGVQGNLEITYLPT 67
      11 GLLPFLARSGSEVNGSQAVCPETLNGLSVTGDAENQVOTLYKLYERCEVWNGNLEIYVLTGH 70
Qy      68 NASLSFLQDIOEYQGYVLIAHQVROVPLQRLIRVGRTOLEFEDVYALAVLDNGLPNNTT 127
      71 NADLSFLQWIREYTGVLVVMNEFSTLPLNLRKVVKTGYVDGKFALFVM---LNYNT 125
Qy      128 PVTGASPGGLRELQLSLTLEILKGVLIQRPOLCYODTILMKDIFHKNNQLALLTIDTN 187
      126 ---NSSHALRQLRLQVLTLEILSGVYIEKNDLCHMDITDMDIVADRQ---AEIYKD 178
Qy      188 RSRACHPCSPMCKSGRCWSESSDCSLTRTVAGGC-ARCKGRLPOLYIKANSKFGITE 246
      179 NGRSCPPECHVCKG-RCWGGSSEDCQTLTKTICAPQCNHCFSPNPVQCHDCACGCGSG 237
Qy      247 LKHSDDLCLLHFHNSGICELHCPALVTYNTDTESMPNEGRYTFGASCYTAAPRYNLTST 306
      238 PDDTDFACRHRFDSDACVPRCPQPLVYNKLTQLEBNPTKYOYGGVCAASCPHNFV-V 296
Qy      307 DVGSCTLVCPNLHNEVTAEDGTORCEKSCPCARVCGIGMEHLREVRATYSANIOEFAQ 366
      297 DOTS CVRACPRPDMMEVD-KNGLKMCEPCGGLCPKACBGSGG--SRFQYVDSNNINGFVN 353
Qy      367 CKKIFGSLAFLPESPFDGPASNTAPLOEOLQVFETLEITGYLYISAMPDSLPDLVSFQ 426
      354 CTKLIGNLDFLLTGLNGDPMHKIPALDPEKLVNFRVREITGYLNIQSWPRPHNFSVFS 413
Qy      427 NLQVIRGRILHNGAYS-LTLQGLIGISMLGRSIRELSGSLALIHNTHLCEVHTVPMDDL 485

```

```

Db      414 NLTTIGRSILNKGFSLLIMKLNVTSLGFSRLKEISAGRIYISANROLCVHHSIMNTKV
Oy      486 FRNPHOALLHTA-NRPEDCEVGEGLACHOLCARGHCMGPGPTQCYNCSOPLRGQECVEEC
Db      474 LRGTLEERLDIKHNRPRDCVAEGKVCPLCSSGCGMCPGQCLSCNYSRSGVCYTHC
Oy      545 RVLOGLPREVYNARHCLPCHPECCOPBONSVCFCPEADQVCAAHYKDPFCVACRPSGV
Db      534 NFLNGEPREFAHAEACFCSCHEPCQMEGTATCGSGSDTCAQCAHFRDGPCHVSSCPHGV
Oy      605 KPDLSYMPIWKFPDEEGACQCPINCTHSCVDLDKCGPAEGR-----SPLTISYAVNG
Db      594 LG--AKGPIYKYPDVQNECRCHENCCTCGCKGPELDGLGTLVLIGKTHLTMAITLVAG
Oy      661 ILVVVLGVFVIGILIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNOQMRLKETE
Db      652 --LVVIFMVLGGLTFLYMGRIQNGRMRVLERGESIEPLDPS-EKANKYLARFKETE
Oy      720 LRKXVLGSGAFGTYVKGIMI PDGENVKI PVAIKVLRNTSPKANKELIDEAYYVAGVS
Db      709 LRKLKVLGSGVFGTVHKGVMIPEGESIKIPVCIKVIEBKSGRQSFQAVTDHMLAIGSLDH
Oy      780 PYVRRLICLTSTVQVLTQMLPYCCLLDHYRENRGRIGSDLLMCMQIAKMSYLEDV
Db      769 AHIVRLGLCGSSSLQVLTQYLPGLSLDHYRQHRGALGPOLLNMVGQIAKMYLYEBH
Oy      840 RLTVRDLAARVNLVKS PMHVKITDEGLARLLDIDETEYHADGQKPIKMMALLESILRRRF
Db      829 GMVRNLAAARVNLKSPSOVQVADPGVADLLRPDDKOLLYSEAKPIKMMALLESILHFKY
Oy      900 THQSDVMSYGVTVLWELMTFGAKPYDGIAPAREIPDLLEKGERLPOPPICITIDVYMIWVCW
Db      889 THQSDVMSYGVTVLWELMTFGAEPYAGRLAEVPLDEKGERLAQPOICTIDVYMWVCW
Oy      960 MIDECRRREFLYSEFSRMARDPQRFVYIQNEDLGRA---SPLDSTYRSLLEDDMGD
Db      949 MIDENIRTEKELANEFTRMARDPRLYLVIKRES-GPGIAPGPEPHGLTNKKLEVELEP
Oy      1017 LVDAEYLVPOQGFCCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEP-SEEARSPPL
Db      1008 ELDDLDLEAED-----NLATTLTGLSALSPVOTLNRPGSGLL
Oy      1076 APSGAGSDVDGDLKMGAAKGLSLPTHN-PSPLQRYSEDPVPLP-----SETDGYV
Db      1049 SPSSGY-MPMNQNLGEBSCQSAVSGSSERCPRPVSLH-----PMRGLASSESGHV
Oy      1129 A-----PLTSGPQE-----YVNOVDVAPQPPSPREG-----L
Db      1102 TGSEAELOEKVSMCRSRSRSPRPRGDSAYHSQHSLLTPTVPLSPGLEEDVNGYV
Oy      1158 PAARPAATLERAKTLSP-GKNGV-----KDVAFAGAVENPEYLTPOGGAAPQHP
Db      1162 PDTHLKGTSPSRREGTLSSVGLSVLTGTEDEED-----EVEYEMRRRRHSP-PHP
Oy      1211 PAPSAPADNLTYWD-----QDPERGAPPTFKGTPTAENEYLT
Db      1213 RPSLSLELGYEMDVGSDLSASLSTGSCPLHPVPIPTAGTTPDEDEYEM

```

RESULT 10

```

Jc4387
Epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: Jc4387
R:Helixer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: Jc4387; MUID:96096535; PMID:8522190
A:Accession: Jc4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

```

```

A:Cross-references: GB:U29339; NID:9915389; PID:9915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residu
C:Comment: This protein is a functional heregulin receptor that transduces signals to
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homc
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)
Query Match 33.3%; Score 2260.5; DB 2; Length 1339;
Best Local Similarity 40.1%; Pred. No. 2e-89;
Matches 515; Conservative 170; Mismatches 443; Indels 155; Gaps 34;
Oy      3 LALCRMGCLLALPPGAA---STQVCTGDMKRLRPASPEHLDMLRHLVQCGOVGN
Db      7 LQVLC---FLSLARGSEKNSQAVCEPTGLANGSLVTDADNOYOTLYKLEKEEVNQN
Oy      60 LELTYLPTNLSLFLQDIQEVQYVLAHNOVROVPLQRLIRVRCQTLFEDNYALAVLDN
Db      63 LEIVLTGHMADSLFLQWIREVTAVYLVAMNEFSVLPLENLAVNGTQYVDGKFAIFVM--
Oy      120 GDPANNTPTVGAAPGGLRELQSLSTEILKGVLIQNPOLCTODITLMKDIFHKNNOL
Db      121 ---LNTYNT---NSSHALRQLKFTOLTLEISGVYIEKNDKLCMDITIDMRDVIYVR--
Oy      180 ALTIDITRSRACHPCSPMKSGRCSWGSSEDCSLPTFTVACGC-ARCKGPELQYIAN
Db      171 GAELVKNKANCPCPCEVCKG-RCKGPGDDCOLITITCAPQCNCRFGENPQOCHD
Oy      239 SKFIQITELKASDCLACLFHNSGICELCPALTYNTDTFESMNPREGRTFGASCYTA
Db      230 ECAGCGSGPODTCFACRFNDSGACVPRCEPLVYNNLTQLENNPHTKXYOGGVCAAS
Oy      239 CPYNYLTSDVSCTLVCPYLNQEVTAEDGTORCEKSKFPCARVCYGLGMEHLREVRATVS
Db      290 CPNHFV-VDQFCVRACPPDMKVD-KHGLMKCEPCGGLCPKAEGETGSG--SRVQYDVS
Oy      359 ANIQFAGCKKIFSLAFLPESFDGPDASNTAPLOPELOVFEETLEETLYGLVYISAMPDS
Db      346 SNIDGFVNCTKILGNLDPLTGLAVDPMHKIPALDPEKLVNFRVRELTGLYNTIOSWPRH
Oy      419 LPDLVSFQNLQVIRGLIHNGAYS-LTLOGIGISMLGRLSRLRELSGLALIHNTHLGCFV
Db      406 MHNFSVFSNLTIGRSLYNRGFSLLIMKLNVTSLGFSRLKEISAGRVYISANQOLCYH
Oy      478 HTVPMQDLFRNPHOALLHTA-NRPEDCEVGEGLACHOLCARGHCMGPGPTQCYNCSOPLR
Db      466 HSLMTBLLRGPSEERLDIKYDRPLGECLEAGKCDPLCSSGCGMCPGQCLSCNYSR
Oy      537 GQECVCECRVYLOGLPREVYNARHCLPCHPECCOPBONSVCFCPEADQVCAAHYKDPFC
Db      526 EGVCTVTHCNFLQGRPREVHAQCFSCHPCLPMEGSTYNGSSDCAARCARFRDGHG
Oy      597 VARCPGVPKPDLSYMPIWKFPDEEGACQCPINCTHSC--VDLDKCGPAEQRASPLTSI
Db      586 VNSCPFIIG--AGPIYKYPDAQNECRCHENCCTCGCKGPELDGLGTLVLAHMLSKPHLV
Oy      655 VSAVVGILLVYVGVIGILIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNOQMRLKETE
Db      644 IAVTVG--LAVIIMLIGSFLYMGRIQNGRMRVLERGESIEPLDPS-EKANKYLAR
Oy      714 ILKETELKXVVLGSGAFGTYVKGIMI PDGENVKI PVAIKVLRNTSPKANKELIDEAYV
Db      701 IFKETELKXVVLGSGVFGTVHKGVMIPEGESIKIPVCIKVIEBKSGRQSFQAVTDHMLA
Oy      774 MAGVSPYVSRLLGLICTSTVQVLTQMLPYCCLLDHYRENRGRIGSDLLMCMQIAKMS

```


Query Match 25.1%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 5.8e-66;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

587 CAHYDPPPCVACRCSPGPDLSYMPIMKFPDEBACQPCPINCHSCYVDLDDKCPRAQ 646
3 CAHFDGPHCVACVACGAVLGENDTL-VRYADANAVCOLCHPCRCGCGPGLCECP--- 58
647 RASPLTISVAVV-GILLVVLGVVGLIKRROOKIRKYTRMRLLQETLEVEPLTPSGA 705
59 NSGKTPSIAGVGVGLCLLVVGLIGLYLRRR-HIVRKTIRLLQETLEVEPLTPSGE 117
706 MPNQAOIRILKETELRKVYVLSGAFGTYYKGIWIPDGENVKI PVAIKVLRNENTSPKANK 765
118 APNQAHRLILKETEFKVKVLSGAFGTYYKGLIPEGEVKI PVAIKELREATSPKANK 177
766 EILDDAAYVAGVGSYVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRNRRGLSGDNLN 825
178 EILDDAAYVAGVGSYVSRLLGICLTSTVQLVTQLMPYGCCLLDYIREHKNIGSQYLLN 237
826 CMQIAKGSYLEVDVLRHDLAARVLYKSPNHVITDGLARLLDIDETEHADGKVP 885
228 CVQIAKGMNLYLEERLVHDLAARVLYKTPQHVKITDGLAKLIGADEKEVTHAEGKVP 297
886 IKMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLKGERLPQRP 945
298 IKMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQRP 357
946 ICTIDVYMIWVCMWIDSECRPFRELVEFSRMARDPQRFVVIQ-NEDLGASPLDSTF 1004
358 ICTIDVYMIWVCMWIDSECRPFRELVEFSRMARDPQRFVVIQ-NEDLGASPLDSTF 417
1005 YRSLEDDDMGDLVDAEYLVPOQGFPCDPPAPGAGMWHHRSSSTFGSGDITLGL 1064
418 YRTLMEEEDMEIVDAEYLVPHQGF-----NSPST----- 449
1065 PSEERAPSPPL-----APSEGAGSDVFDGLGMAKGLQSLPTHPSPQYSEDPVP 1119
450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVQYSSDPTGN 495
1120 LPSET--DGVAAPLTCSPQPEYVNOQDVRRPQPSBREGPLPAARAGATLEAKTSLSECK 1177
496 FLEESIDGFL-----PAPETVNO--LMPKKPSTAM----- 524
1178 NGVVDKVPFAF-----GGAIVENPEYLPQSGAQPQHPAPSPAFD 1218
525 --VQGIYNFISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSP LAKTYFE 574
1219 NLVYWDQDPPERGAPSTFKGTPTAENPEY 1248
575 SSPVWIOGNNQ-----INLDNPDY 594

RESULT 13
S35745
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
Submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL.X12707
C:Genetics:
A:Gene: erdB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F
F.143-400/Domain: protein kinase homology <KIN>
F.143-151/Region: protein kinase ATP-binding motif
F.170/Active site: Lys #status predicted

Query Match 24.3%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.3e-63;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

578 GREADQCAAHYKDPPECVACRCSPGPDLSYMPIMKFPDEBACQPCPINCHSCYVDL 637
1 GP--DHCKCAHFDGPHCVACVACGAVLGENDTL-VKRYADANAVCOLCHPCRCGCKP 57
638 DDGCPAQRASPLTISVAVV-GILLVVLGVVGLIKRROOKIRKYTRMRLLQETLE 696
58 GLEGP---NSGKTPSIAGVGVGLCLLVVGLIGLYLRRR-HIVRKTIRLLQETLE 113
697 VEPLTPSGAMPQAOIRILKETELRKVYVLSGAFGTYYKGIWIPDGENVKI PVAIKVLR 756
114 VEPLTPSGAMPQAOIRILKETEFKVKVLSGAFGTYYKGLIPEGEVKI PVAIKELR 173
757 ENTSPKANKEILDEAYVAGVGSYVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRNRR 816
174 EATSPKANKEILDEAYVAGVGSYVSRLLGICLTSTVQLVTQLMPYGCCLLDYIREHKN 233
817 LGSQDLWVCMQIAKGSYLEVDVLRHDLAARVLYKSPNHVITDGLARLLDIDETE 876
234 IGSQDLWVCMQIAKGMNLYLEERLVHDLAARVLYKTPQHVKITDGLAKLIGADEKE 293
877 YHADGKVPIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLL 936
294 YHADGKVPIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLE 353
937 KGERLPQRPICITIDVYMIWVCMWIDSECRPFRELVEFSRMARDPQRFVVIQ-NEDLG 995
354 KGERLPQRPICITIDVYMIWVCMWIDSECRPFRELVEFSRMARDPQRFVVIQ-NEDLG 413
996 PASPLDSTFYRSLEDDDMGDLVDAEYLVPOQGFPCDPPAPGAGMWHHRSSSTSG 1055
414 LPSPTDSKFYRTLMEEEDMEIVDAEYLVPHQGF-----NSPST--- 454
1056 GQDLTLGLEPSEERAPSPPL-----APSEGAGSDVFDGLGMAKGLQSLPTHPSPQ 1110
455 -----SRTPLLSSLSATSNNSATNCIDKNG-----H----- 481
1111 RYSEDPVPLPSETDGYVAPLTCSPQPEYVNOQDVRRPQPSBREGPLPAARAGAT-LE 1169
482 -----PVREDFGL-----PAPETVNO--LMPKKPSTAM----- 523
1170 AKTISPGKGVKVPFAFAGAVENPEYL 1197
524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14
S00727
K:kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Janzen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erdB gene of avian erythroblastosis virus mut.
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL.X06943
C:Genetics:
A:Gene: erdB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F.135-400/Domain: protein kinase homology <KIN>
F.143-151/Region: protein kinase ATP-binding motif

Query Match 24.2%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 2.6e-63;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-210-224-12

Perfect score: 6789
Sequence: 1 MELALCRMGILLALLPPGA.....TFKGTPTAENPEYIGLDVVP 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6698	98.7	1255	1	ERB2_HUMAN
2	5886	86.7	1257	1	ERB2_RAT
3	5876.5	86.6	1254	1	ERB2_MESAU
4	3074	45.3	1210	1	EGFR_HUMAN
5	3053	45.0	1210	1	EGFR_MOUSE
6	2909.5	42.9	1308	1	ERB4_HUMAN
7	2890	42.6	1308	1	ERB4_MOUSE
8	2625.5	38.7	1167	1	XMRK_XIPMA
9	2354.5	34.7	1167	1	ERB3_HUMAN
10	2283.5	33.6	1339	1	ERB3_RAT
11	1891	27.9	1426	1	EGFR_DROME
12	1749.5	25.8	634	1	ERBB_ALV
13	1703	25.1	604	1	ERBB_AVIER
14	1630	24.0	540	1	ERBB_AVIER
15	1532	22.6	703	1	EGFR_CHICK
16	1255	18.5	1323	1	LR23_CAEL
17	1142.5	16.8	245	1	ERB2_MOUSE
18	675	9.9	1363	1	ILPR_BRALA
19	673	9.9	1382	1	INSR_HUMAN
20	666	9.8	1383	1	INSR_RAT
21	665.5	9.8	1372	1	INSR_MOUSE
22	665	9.8	1297	1	IRR_HUMAN
23	662	9.8	1300	1	IRR_MOUSE
24	662	9.8	1607	1	MIRP_LYMST
25	659.5	9.7	1300	1	IRR_CAVPO
26	642	9.5	1477	1	HTK7_HYDAT
27	603	8.9	1367	1	IGIR_HUMAN
28	596	8.8	1373	1	IGIR_MOUSE
29	595	8.8	987	1	EPB4_HUMAN
30	592.5	8.7	1370	1	IGIR_RAT
31	589.5	8.7	977	1	EPB2_MOUSE
32	589	8.7	987	1	EPB4_MOUSE
33	588	8.7	1114	1	RET_HUMAN

34	588	8.7	1390	1	INSR_AEDAE
35	588	8.7	2146	1	INSR_DROME
36	584.5	8.6	976	1	EPB2_HUMAN
37	582.5	8.6	984	1	EPB1_RAT
38	580.5	8.6	984	1	EPB1_CHICK
39	576.5	8.5	984	1	EPB1_HUMAN
40	573.5	8.4	1053	1	FAK1_CHICK
41	569	8.4	1068	1	FAK1_XENLA
42	567	8.4	757	1	HT16_HYDAT
43	567	8.4	902	1	EPB8_XENLA
44	566.5	8.3	985	1	EPB4_XENLA
45	563	8.3	1052	1	FAK1_MOUSE

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD: PRT: 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).
DE ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; Pubmed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=86070181; Pubmed=2999974;
RX Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
[3]
RN SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; Pubmed=2995967;
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;
RT "A c-erbB-2-related protooncogene, c-erbB-2, is distinct from the human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
[4]
RN VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; Pubmed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization.";
RL Genomics 15:426-429(1993).
-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).
-1- SUBCELLULAR LOCATION: Type I membrane protein.

FT	DISULFID	563	576	BY SIMILARITY.
FT <td>DISULFID</td> <td>567</td> <td>584</td> <td>BY SIMILARITY.</td>	DISULFID	567	584	BY SIMILARITY.
FT <td>DISULFID</td> <td>587</td> <td>596</td> <td>BY SIMILARITY.</td>	DISULFID	587	596	BY SIMILARITY.
FT <td>DISULFID</td> <td>600</td> <td>623</td> <td>BY SIMILARITY.</td>	DISULFID	600	623	BY SIMILARITY.
FT <td>DISULFID</td> <td>626</td> <td>634</td> <td>BY SIMILARITY.</td>	DISULFID	626	634	BY SIMILARITY.
FT <td>DISULFID</td> <td>630</td> <td>642</td> <td>BY SIMILARITY.</td>	DISULFID	630	642	BY SIMILARITY.
FT <td>MOD_RES</td> <td>1139</td> <td>1139</td> <td>PHOSPHORYLATION (AUTO-) (BY SIMILARITY).</td>	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT <td>MOD_RES</td> <td>1248</td> <td>1248</td> <td>PHOSPHORYLATION (AUTO-) (BY SIMILARITY).</td>	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT <td>CARBOHYD</td> <td>68</td> <td>68</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>CARBOHYD</td> <td>124</td> <td>124</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>CARBOHYD</td> <td>187</td> <td>187</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>CARBOHYD</td> <td>259</td> <td>259</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>CARBOHYD</td> <td>530</td> <td>530</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>CARBOHYD</td> <td>571</td> <td>571</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	571	571	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>CARBOHYD</td> <td>629</td> <td>629</td> <td>N-LINKED (GLCNAC. . .) (POTENTIAL).</td>	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT <td>VARIANT</td> <td>654</td> <td>654</td> <td>I -> V.</td>	VARIANT	654	654	I -> V.
FT <td>VARIANT</td> <td>655</td> <td>655</td> <td>/FTID=VAR_004077.</td>	VARIANT	655	655	/FTID=VAR_004077.
FT <td>VARIANT</td> <td>655</td> <td>655</td> <td>I -> V.</td>	VARIANT	655	655	I -> V.
FT <td>CONFLICT</td> <td>1170</td> <td>1170</td> <td>/FTID=VAR_004078.</td>	CONFLICT	1170	1170	/FTID=VAR_004078.
FT <td>CONFLICT</td> <td>1255</td> <td>1255</td> <td>P -> A (IN REF. 2).</td>	CONFLICT	1255	1255	P -> A (IN REF. 2).
SEQ <td>SEQUENCE</td> <td>1255</td> <td>137909</td> <td>MM; 39E9DDDA04DCE962 CRC64;</td>	SEQUENCE	1255	137909	MM; 39E9DDDA04DCE962 CRC64;
Query March				
Best Local Similarity		98.7%;	Score 6698; DB 1; Length 1255;	
Matches 1240; Conservative		98.8%;	Pred. No. 0;	
			1; Mismatches 14; Indels 0; Gaps 0;	

Qy	1	MEI ¹ ALACRMGILLALALPEGAAS ² TQVCTGTD ³ IMK ⁴ LR ⁵ PLASPE ⁶ TH ⁷ DL ⁸ ML ⁹ RL ¹⁰ YL ¹¹ QGVCO ¹² VQNL	60
Dh	1	MEI ¹ ALACRMGILLALALPEGAAS ² TQVCTGTD ³ IMK ⁴ LR ⁵ PLASPE ⁶ TH ⁷ DL ⁸ ML ⁹ RL ¹⁰ YL ¹¹ QGVCO ¹² VQNL	60
Qy	61	ELTYLPT ¹ NASL ² SFLQD ³ IQEVQ ⁴ YVL ⁵ AHNOV ⁶ ROV ⁷ PLQRL ⁸ IVRG ⁹ TO ¹⁰ LFEDNY ¹¹ ALAV ¹² LDNG	120
Dh	61	ELTYLPT ¹ NASL ² SFLQD ³ IQEVQ ⁴ YVL ⁵ AHNOV ⁶ ROV ⁷ PLQRL ⁸ IVRG ⁹ TO ¹⁰ LFEDNY ¹¹ ALAV ¹² LDNG	120
Qy	121	DPLANN ¹ TPV ² TGASGG ³ IRELQ ⁴ LSLT ⁵ ILKGCV ⁶ LIQ ⁷ NPOLCYD ⁸ DT ⁹ ILMKD ¹⁰ IFHKNQ ¹¹ LA	180
Dh	121	DPLANN ¹ TPV ² TGASGG ³ IRELQ ⁴ LSLT ⁵ ILKGCV ⁶ LIQ ⁷ NPOLCYD ⁸ DT ⁹ ILMKD ¹⁰ IFHKNQ ¹¹ LA	180
Qy	181	LTL ¹ LD ² TRSRACH ³ PCSPCKGSR ⁴ CMGSESS ⁵ EDCS ⁶ L ⁷ TRTVACG ⁸ CARCK ⁹ PL ¹⁰ YIKANSK	240
Dh	181	LTL ¹ LD ² TRSRACH ³ PCSPCKGSR ⁴ CMGSESS ⁵ EDCS ⁶ L ⁷ TRTVACG ⁸ CARCK ⁹ PL ¹⁰ YIKANSK	240
Qy	241	FIGITEL ¹ KHS ² DCLAC ³ LHFNHSG ⁴ ICELHCPALVT ⁵ YNTD ⁶ FESMPN ⁷ BEGR ⁸ YTFGASCV ⁹ TACP	300
Dh	241	AAGGTG ¹ KHSDCLAC ² LHFNHSG ³ ICELHCPALVT ⁴ YNTD ⁵ FESMPN ⁶ BEGR ⁷ YTFGASCV ⁸ TACP	300
Qy	301	YNYLST ¹ DVGSCT ² LYCP ³ LHNOE ⁴ VT ⁵ AEDGT ⁶ QRCCKSK ⁷ CAVCYGLGMEHL ⁸ REVR ⁹ AVT ¹⁰ TSAN	360
Dh	301	YNYLST ¹ DVGSCT ² LYCP ³ LHNOE ⁴ VT ⁵ AEDGT ⁶ QRCCKSK ⁷ CAVCYGLGMEHL ⁸ REVR ⁹ AVT ¹⁰ TSAN	360
Qy	361	IOEPAGCK ¹ K ² FGSLA ³ LPES ⁴ FDGP ⁵ ASNT ⁶ PL ⁷ POEOL ⁸ OVET ⁹ LEIT ¹⁰ EGY ¹¹ YSAMP ¹² SLP	420
Dh	361	IOEPAGCK ¹ K ² FGSLA ³ LPES ⁴ FDGP ⁵ ASNT ⁶ PL ⁷ POEOL ⁸ OVET ⁹ LEIT ¹⁰ EGY ¹¹ YSAMP ¹² SLP	420
Qy	421	DL ¹ SVFQ ² LYIRGR ³ ILHNGA ⁴ YS ⁵ LT ⁶ LQIG ⁷ ISW ⁸ LGR ⁹ LSREL ¹⁰ SGSL ¹¹ ALIH ¹² NTH ¹³ LCF ¹⁴ HTV	480
Dh	421	DL ¹ SVFQ ² LYIRGR ³ ILHNGA ⁴ YS ⁵ LT ⁶ LQIG ⁷ ISW ⁸ LGR ⁹ LSREL ¹⁰ SGSL ¹¹ ALIH ¹² NTH ¹³ LCF ¹⁴ HTV	480
Qy	481	PMDL ¹ FNP ² HQ ³ AL ⁴ LHT ⁵ ANR ⁶ PEDEC ⁷ VEG ⁸ GLACH ⁹ OL ¹⁰ CARG ¹¹ HCW ¹² GP ¹³ Q ¹⁴ CVNCS ¹⁵ OF ¹⁶ LRQ ¹⁷ QBC	540
Dh	481	PMDL ¹ FNP ² HQ ³ AL ⁴ LHT ⁵ ANR ⁶ PEDEC ⁷ VEG ⁸ GLACH ⁹ OL ¹⁰ CARG ¹¹ HCW ¹² GP ¹³ Q ¹⁴ CVNCS ¹⁵ OF ¹⁶ LRQ ¹⁷ QBC	540
Qy	541	VEEB ¹ RV ² QGLPRE ³ EVNAR ⁴ HL ⁵ PC ⁶ HPE ⁷ QOP ⁸ NGS ⁹ YTC ¹⁰ GP ¹¹ RA ¹² QCVAC ¹³ AH ¹⁴ KY ¹⁵ DP ¹⁶ FC ¹⁷ YARC	600
Dh	541	VEEB ¹ RV ² QGLPRE ³ EVNAR ⁴ HL ⁵ PC ⁶ HPE ⁷ QOP ⁸ NGS ⁹ YTC ¹⁰ GP ¹¹ RA ¹² QCVAC ¹³ AH ¹⁴ KY ¹⁵ DP ¹⁶ FC ¹⁷ YARC	600
Qy	601	PSGK ¹ PDL ² SL ³ MP ⁴ IMK ⁵ PFDEGA ⁶ COP ⁷ CE ⁸ INT ⁹ HS ¹⁰ CV ¹¹ DD ¹² GC ¹³ PA ¹⁴ QAS ¹⁵ PL ¹⁶ TS ¹⁷ IVSA ¹⁸ VG	660
Dh	601	PSGK ¹ PDL ² SL ³ MP ⁴ IMK ⁵ PFDEGA ⁶ COP ⁷ CE ⁸ INT ⁹ HS ¹⁰ CV ¹¹ DD ¹² GC ¹³ PA ¹⁴ QAS ¹⁵ PL ¹⁶ TS ¹⁷ IVSA ¹⁸ VG	660
Qy	661	ILL ¹ VVL ² GV ³ FG ⁴ IL ⁵ IKRQ ⁶ K ⁷ IRY ⁸ TM ⁹ RL ¹⁰ QET ¹¹ VEL ¹² VP ¹³ SGAMP ¹⁴ QA ¹⁵ OM ¹⁶ RL ¹⁷ KET ¹⁸ EL	720
Dh	661	ILL ¹ VVL ² GV ³ FG ⁴ IL ⁵ IKRQ ⁶ K ⁷ IRY ⁸ TM ⁹ RL ¹⁰ QET ¹¹ VEL ¹² VP ¹³ SGAMP ¹⁴ QA ¹⁵ OM ¹⁶ RL ¹⁷ KET ¹⁸ EL	720

Oy		721	RKVAVLSGAGCTAYKKGWIDGENVVKI.PVAIKVLRENTSPRANKKEIIDENYVMAGVSP	780
Dd		721	RKVRLSGAGAGTGYKKGIWIDGENVVKIPVAIKVLRENTSPRANKKEIIDENYVMAGVSP	780
Oy		781	YVSRLGLCLSTVOLVTQLMPYGCGLDHAVERNRLGSODPLNMCQIAIGMSYLEDV	840
Dd		781	YVSRLGLCLSTVOLVTQLMPPYGCGLDHAVERNRLGSODPLNMCQIAIGMSYLEDV	840
Oy		841	LVRHDLAARNLVKSPNHVKITDFGLARLDIDETEHADGGKVPDKMALESILRRFT	900
Dd		841	LVRHDLAARNLVKSPNHVKITDFGLARLDIDETEHADGGKVPDKMALESILRRFT	900
Oy		901	HOSDWMSGVTVWELMTFGAPRYDICI.PAREIPDLLEKEERLPPOPCTIDVYMMIVCKM	960
Dd		901	HOSDWMSGVTVWELMTFGAPRYDICI.PAREIPDLLEKEERLPPOPCTIDVYMMIVCKM	960
Oy		961	IDSECRPRFRRELVEFSFSSMARDDPORFVVIQNEIDLGPASPLDSTFYRSLLIEDDMGDLYDA	1020
Dd		961	IDSECRPRFRRELVEFSFSSMARDDPORFVVIQNEIDLGPASPLDSTFYRSLLIEDDMGDLYDA	1020
Oy		1021	EELYVPOQGFFCCPDPAFAGCMVHHRRSSSTRSGGDLTYGLEPSEBEAFRSLAPSEG	1080
Dd		1021	EELYVPOQGFFCCPDPAFAGCMVHHRRSSSTRSGGDLTYGLEPSEBEAFRSLAPSEG	1080
Oy		1081	AGSDVFDDDLGMGAAGLQSLPTPHDSPLOYSEDPTVPLSETDGVYAPLTCSPPEPYV	1140
Dd		1081	AGSDVFDDDLGMGAAGLQSLPTPHDSPLOYSEDPTVPLSETDGVYAPLTCSPPEPYV	1140
Oy		1141	NQPDVRPOPSPPRECEPLPAABPAGATLERAKTLSPGNKYVDYFAFGAENVPEYLTPQ	1200
Dd		1141	NQPDVRPOPSPPRECEPLPAABPAGATLERAKTLSPGNKYVDYFAFGAENVPEYLTPQ	1200
Oy		1201	GGAAPQHPPPPAFSPAFPNLYYMDODPPERGAPSTFGGTPTAENPEYLGLDVPY	1255
Dd		1201	GGAAPQHPPPPAFSPAFPNLYYMDODPPERGAPSTFGGTPTAENPEYLGLDVPY	1255

RESULT 2

ERR2_RAT		STANDARD:	PRT: 1257 AA.
AC	P06494;		
DT	01-JAN-1968 (Rel. 06, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor Protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112) (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).		
DN	ERRB2 OR NEU.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxId=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Neuroblastoma;		
RX	MEDLINE=86118662; PubMed=3945311;		
RA	Bargmann C.I., Hung M.-C., Weinberg R.A.;		
RT	"the neu oncogene encodes an epidermal growth factor receptor-related protein.";		
RL	Nature 319:226-230(1986).		
RN	[2]		
RP	SEQUENCE OF 852-905 FROM N.A.		
RC	TISSUE=Sciatic nerve;		
RX	MEDLINE=91222560; PubMed=2025425;		
RA	Lai C., Lemke G.;		
RT	"An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";		
RL	Neuron 6:691-704(1991).		
RN	[3]		
RP	STRUCTURE BY NMR OF 650-668.		
RX	MEDLINE=92155181; PubMed=1346763;		
RA	Gullick W.J., Bottomley A.C., Lotfs F.U., Doak D.G., Mulvey D.,		

[illegible]

```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MM; 6129264583011402 CRC64;

Query Match 86.7%; Score 5886; DB 1; Length 1257;
Best Local Similarity 86.8%; Pred. No. 3,3e-311;
Matches 1091; Conservative 50; Mismatches 114; Indels 2; Gaps 2;

OY 1 MELALCRWGLLALLPGASQVCTGDMKRLPASPEHLDMLRHLYOGCQVQGNL 60
D 1 MELAACRWGFLALLPGIAGTQCTGDMKRLPASPEHLDMLRHLYOGCQVQGNL 60
OY 61 ELTYLPFNASLFLQDIOEVQGYVLAHNOVROVPLQRLRIVRGTLFEDNYALAVLDNR 120
D 61 ELTYLPANASLFLQDIOEVQGYVLAHNOVKRPLQRLRIVRGTLFEDNYALAVLDNR 120
OY 121 DPLANTPTV-GASPGGLRELQRLSTELLKGVLIQNRPOLCYQDTILMKDIFHKNNOL 179
D 121 DPLANTPTV-GASPGGLRELQRLSTELLKGVLIQNRPOLCYQDTILMKDIFHKNNOL 179
OY 121 DPLANTPTV-GASPGGLRELQRLSTELLKGVLIQNRPOLCYQDTILMKDIFHKNNOL 180
D 121 DPLANTPTV-GASPGGLRELQRLSTELLKGVLIQNRPOLCYQDTILMKDIFHKNNOL 180
OY 180 ALFLIDNRRACHPCSPMKGSRGSGESSDQSLTRTVGAGGACRCKGFLPOYIKANS 239
D 180 ALFLIDNRRACHPCSPMKGSRGSGESSDQSLTRTVGAGGACRCKGFLPOYIKANS 239
OY 181 APVDIDNRRACHPCSPMKGSRGSGESSDQSLTRTVGAGGACRCKGFLPOYIKANS 240
D 181 APVDIDNRRACHPCSPMKGSRGSGESSDQSLTRTVGAGGACRCKGFLPOYIKANS 240
OY 240 KFIGITELKSDCLACHFNHSGICELHCPALVTYNTDTFSSMPREBYTFGASCVTAC 299
D 240 KFIGITELKSDCLACHFNHSGICELHCPALVTYNTDTFSSMPREBYTFGASCVTAC 299
OY 241 CAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFSSMPREBYTFGASCVTAC 300
D 241 CAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFSSMPREBYTFGASCVTAC 300
OY 300 PYNLYSTDVGSCTVLCPLAHNOVTRAEDETCRCKSCARCYGLGMEHLREVAVTSA 359
D 300 PYNLYSTDVGSCTVLCPLAHNOVTRAEDETCRCKSCARCYGLGMEHLREVAVTSA 359
OY 301 PYNLYSTDVGSCTVLCPLAHNOVTRAEDETCRCKSCARCYGLGMEHLREVAVTSA 360
D 301 PYNLYSTDVGSCTVLCPLAHNOVTRAEDETCRCKSCARCYGLGMEHLREVAVTSA 360
OY 360 NIOFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLQVFEITLITGYLYISAMPDL 419
D 360 NIOFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLQVFEITLITGYLYISAMPDL 419
OY 361 NIOFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLQVFEITLITGYLYISAMPDL 420
D 361 NIOFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLQVFEITLITGYLYISAMPDL 420
OY 420 PDLSTFQNLQVIRGRILHNGAYSLTLQGLISWLGSLRSLREISGLALIHNTHLCPVHT 479
D 420 PDLSTFQNLQVIRGRILHNGAYSLTLQGLISWLGSLRSLREISGLALIHNTHLCPVHT 479
OY 421 RDLSTFQNLQVIRGRILHNGAYSLTLQGLISWLGSLRSLREISGLALIHNTHLCPVHT 480
D 421 RDLSTFQNLQVIRGRILHNGAYSLTLQGLISWLGSLRSLREISGLALIHNTHLCPVHT 480
OY 480 VPMOULFRNPHOALLHTANRPEDE-CVSGGLACHOU-CARHGMGPPPTQVNCSCPLRQ 538
D 480 VPMOULFRNPHOALLHTANRPEDE-CVSGGLACHOU-CARHGMGPPPTQVNCSCPLRQ 538
OY 481 VPMOULFRNPHOALLHTANRPEDE-CVSGGLACHOU-CARHGMGPPPTQVNCSCPLRQ 540
D 481 VPMOULFRNPHOALLHTANRPEDE-CVSGGLACHOU-CARHGMGPPPTQVNCSCPLRQ 540
OY 539 ECVBEERVLQGLPREVYNARHCLPCHPECOPOGNSVTCFPGPEADQCVACAHYKOPPCVA 598
D 539 ECVBEERVLQGLPREVYNARHCLPCHPECOPOGNSVTCFPGPEADQCVACAHYKOPPCVA 598
OY 541 ECVBEERVLQGLPREVYNARHCLPCHPECOPOGNSVTCFPGPEADQCVACAHYKOPPCVA 600
D 541 ECVBEERVLQGLPREVYNARHCLPCHPECOPOGNSVTCFPGPEADQCVACAHYKOPPCVA 600
OY 599 RCPGSGVPLDLYNPIKFPDEBACOPCPINCHSCVDLDDKCCPAEORASPLTISYSAV 658
D 599 RCPGSGVPLDLYNPIKFPDEBACOPCPINCHSCVDLDDKCCPAEORASPLTISYSAV 658
OY 601 RCPGSGVPLDLYNPIKFPDEBACOPCPINCHSCVDLDDKCCPAEORASPLTISYSAV 660
D 601 RCPGSGVPLDLYNPIKFPDEBACOPCPINCHSCVDLDDKCCPAEORASPLTISYSAV 660
OY 659 VGLLVVLTGAVGILIKRQOKIRKTYMRLLQETELVPLTFSGAMPQAOQRILKET 718
D 659 VGLLVVLTGAVGILIKRQOKIRKTYMRLLQETELVPLTFSGAMPQAOQRILKET 718
OY 661 VGLLVVLTGAVGILIKRQOKIRKTYMRLLQETELVPLTFSGAMPQAOQRILKET 720
D 661 VGLLVVLTGAVGILIKRQOKIRKTYMRLLQETELVPLTFSGAMPQAOQRILKET 720
OY 719 ELKRVKVLGSGAFCTVYKGIWIPDGENVKI PVAIKVLRNTSPRANKELIDEAYVMAVG 778
D 719 ELKRVKVLGSGAFCTVYKGIWIPDGENVKI PVAIKVLRNTSPRANKELIDEAYVMAVG 778
OY 721 ELKRVKVLGSGAFCTVYKGIWIPDGENVKI PVAIKVLRNTSPRANKELIDEAYVMAVG 780
D 721 ELKRVKVLGSGAFCTVYKGIWIPDGENVKI PVAIKVLRNTSPRANKELIDEAYVMAVG 780
OY 779 SPVYSRLIGICLTSTVOLVTQMLPYGCLDHAHRENRGRISQDILLNMCQIAKMSYLED 838
D 779 SPVYSRLIGICLTSTVOLVTQMLPYGCLDHAHRENRGRISQDILLNMCQIAKMSYLED 838

```

```

D 781 SPVYSRLIGICLTSTVOLVTQMLPYGCLDHAHRENRGRISQDILLNMCQIAKMSYLED 840
OY 839 VRLVHRLDAARNVLYKSPNHVKTIDFGARLLDIDETEHADGKVPIKMALESILRR 898
D 841 VRLVHRLDAARNVLYKSPNHVKTIDFGARLLDIDETEHADGKVPIKMALESILRR 900
OY 899 FTHQSDVSYGVTLWELMTFGAKPYDGIIPAREIPOLLEGERLPOPICTIDVMMIMVC 958
D 901 FTHQSDVSYGVTLWELMTFGAKPYDGIIPAREIPOLLEGERLPOPICTIDVMMIMVC 960
OY 959 WMIDSECRPFRELVESEFARMADPQRFVIONEDGAPSLDSTFYRSLLEDDDDMDLV 1018
D 961 WMIDSECRPFRELVESEFARMADPQRFVIONEDGAPSLDSTFYRSLLEDDDDMDLV 1020
OY 1019 DAEEYVLPQGFPCPPAPACAGMVAHRRSSSTRSGGDLTYGLEPSEEARSLAPS 1078
D 1021 DAEEYVLPQGFPCPPAPACAGMVAHRRSSSTRSGGDLTYGLEPSEEARSLAPS 1080
OY 1079 EGAGSDVPDGDLMGAAGLQSLPTHDPSFLQRYSEDPTVPLPSEFDGYVAPLTCSPQPE 1138
D 1081 EGAGSDVPDGDLMGAAGLQSLPTHDPSFLQRYSEDPTVPLPSEFDGYVAPLTCSPQPE 1140
OY 1139 YNQPDPVROPESPREGPLPAARPAATLERAKTSLPGKGVVVDVAFGAVENPEYLT 1198
D 1141 YNQPDPVROPESPREGPLPAARPAATLERAKTSLPGKGVVVDVAFGAVENPEYLT 1200
OY 1199 POGGAAPQHPAPFAPFADNLVYMDQPPERGAPSTFGKTPAENPEYIGLDVVP 1255
D 1201 POGGAAPQHPAPFAPFADNLVYMDQPPERGAPSTFGKTPAENPEYIGLDVVP 1257

RESULT 3
ERB2_MESAU STANDARD; PRT: 1254 AA.
ID ERB2_MESAU
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxId=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ishikawa T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y.,
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE BGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D16295; BAA03801.1; .

DR HSSP: P1362; 1FGK.

DR InterPro: IPR000494; EGFR_L_domain.

DR InterPro: IPR000719; Euk_kinase.

DR InterPro: IPR002174; Furin-like.

DR InterPro: IPR001245; Tyr_kinase.

DR InterPro: IPR004019; YLP_motif.

DR Pfam: PF00069; kinase; 1.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF01030; Recep_L_domain; 2.

DR Pfam: PF02757; YLP; 2.

DR ProDom: PD000001; Euk_kinase; 1.

DR SMART: SM00261; FU; 3.

DR SMART: SM00219; Tyrc; 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.

KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

KW Proto-oncogene; Disease mutation.

FT SIGNAL 1 21

FT CHAIN 22 1254

FT DOMAIN 22 652

FT TRANSSEM 653 675

FT DOMAIN 676 1254

FT DOMAIN 158 368

FT DOMAIN 472 644

FT DOMAIN 720 987

FT NP_BIND 726 734

FT BINDING 753 753

FT ACT_SITE 845 845

FT DISULFID 195 204

FT DISULFID 199 212

FT DISULFID 236 244

FT DISULFID 240 252

FT DISULFID 255 264

FT DISULFID 268 295

FT DISULFID 299 311

FT DISULFID 315 331

FT DISULFID 334 338

FT DISULFID 511 520

FT DISULFID 515 528

FT DISULFID 531 540

FT DISULFID 544 560

FT DISULFID 563 576

FT DISULFID 567 584

FT DISULFID 587 596

FT DISULFID 600 623

FT DISULFID 626 634

FT DISULFID 630 642

FT MOD_RES 1139 1139

FT MOD_RES 1247 1247

FT CARBOHYD 68 68

FT CARBOHYD 125 125

FT CARBOHYD 187 187

FT CARBOHYD 259 259

FT CARBOHYD 530 530

FT CARBOHYD 571 571

FT CARBOHYD 629 629

FT VARIANT 658 658

FT VARIANT 659 659

SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 86.6%; Score 5876.5; DB 1; Length 1254;

Best Local Similarity 86.5%; Pred. No. 1,le-310;

Matches 1086; Conservative 58; Mismatches 110; Indels 1; Gaps 1;

QY 1 MELALATGRMGILLALLPLPGASASTOVCTGTDKRLRLPASPETHLMLRLHLYGGCCVQVQNTL 60

DB 1 MELAAWGMGILLALLPLSGASGTOVCTGTDKRLRLPASPETHLMLRLHLYGGCCVQVQNTL 60

QY 61 ELTYLPTNASTSLFDIDIOEVGVYLAHQVROVPLQRLRIYRGTOLEFEDNALAVLDNG 120

DB 61 ELTYLPAANTLSFLDIDIOEVGVYLAHQVHVPQLRLRIYRGTOLEFEDKALAVLDNR 120

QY 121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLLIQRPOLCYODTILMKDIFHKNOQLA 180

DB 121 DPLDNTTATGRTPGELRELQRLSLTEILKGVLLIRGNPOLCYODTIVLWKDVFRKNOQLA 180

QY 181 LTLIDNRSRACHPCSPCKSGKSGESSEDDOSILTRTCAGGACARCKPLPYITANSK 240

DB 181 PVDIDNRSRACHPCAPACKDNHCKWGSSEDDQTLTAPRAVPAARARLPDCCHEOC 240

QY 241 FIGITELKHSDDLACLFHNSGICELHCPALVTYNTDTFESNPNPEGRTFGASCTYACP 300

DB 241 AAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNPEGRTFGASCTTGP 300

QY 301 YNYLSTDVSGCTLVCPPLHQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360

DB 301 YNYLSTEVSGCTLVCPPLNQEVTAEDGTQRCCKSKSCARVCYGLGMEHLRGARATSAN 360

QY 361 IOEFAGCKKIRGSLAFLESFPGDPAASNTAPLQPEQLQVFELEETGLYISAMPDSL 420

DB 361 IOEFAGCKKIRGSLAFLESFPGDPAASNTAPLQPEQLQVFELEETGLYISAMPDSL 420

QY 421 DLSTVFQNLQVIRGRILHNGAVSLTLOGLGISWGLRSLRELSSGLALIHNNHLCFVHTV 480

DB 421 DLSTVFQNLQVIRGRILHNGAVSLTLOGLGISWGLRSLRELSSGLALIHNNHLCFVHTV 480

QY 481 PWDOLFRNPHOALLHTANRPEDECVBEGILACHQLCARHCHWGPCTQVCNSQFLRGQEC 540

DB 481 PWDOLFRNPHOALLHTANRPEDECVBEGILACHQLCARHCHWGPCTQVCNSQFLRGQEC 540

QY 541 VEECVQVLOGLPREYNAHCHLPCHCECQPNQSVTCFPEADQCAACAHKDPFPCVAC 600

DB 541 VEECVQVLOGLPREYNAHCHLPCHCECQPNQSVTCFPEADQCAACAHKDPFPCVAC 600

QY 601 PSGVPRDLSYMPIMKFPDEBEGACOPCPINCHSCYVDLDKGPAPORASPLSISAVNG 660

DB 601 PSGVPRDLSYMPIMKFPDEBEGACOPCPINCHSCYVDLDKGPAPORASPLSISAVNG 660

QY 661 ILVVVLGIVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKETEL 720

DB 661 ILVVVLGIVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKETEL 720

QY 721 RKVRLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTSPRANKELIDEAVYMGVSGP 780

DB 721 RKVRLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTSPRANKELIDEAVYMGVSGP 780

QY 781 YVSRLLGICLTSTVOLVTQMLPYGCLDHYENRGRGLSODLLNMCQIAKMSYLEDVR 840

DB 781 YVSRLLGICLTSTVOLVTQMLPYGCLDHYENRGRGLSODLLNMCQIAKMSYLEDVR 840

QY 841 LVHRLDAAARNVLYVSPNHTKITDGLARLLDIDETEHADGKVIKXIALESILRRRT 900

DB 841 LVHRLDAAARNVLYVSPNHTKITDGLARLLDIDETEHADGKVIKXIALESILRRRT 900

QY 901 HQSDVMSYGVTVWELMTFGARPYDGIIPAREIPDLLEKEERLPQPICTIDVYMIWKKM 960

DB 901 HQSDVMSYGVTVWELMTFGARPYDGIIPAREIPDLLEKEERLPQPICTIDVYMIWKKM 960

QY 961 IDSECRPRFRLVSEFSMARDDPORFVYIIONEDJGAPSLPSTFRSLLEDMDMDLVDA 1020

DB 961 IDSECRPRFRLVSEFSMARDDPORFVYIIONEDJGAPSLPSTFRSLLEDMDMDLVDA 1020

QY 1021 EBYLVPOGFPFPPAPAGAGMVHRRSSSTRSGGDLTLGLBSEBEAPRSPPLAPSEG 1080

DB 1021 EBYLVPOGFPFPPAPAGAGMVHRRSSSTRSGGDLTLGLBSEBEAPRSPPLAPSEG 1080

QY 1081 AGSDVFPGLDAGMAKGLQSLPTHDPSFLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140

DB 1081 AGSDVFPGLDAGMAKGLQSLPTHDPSFLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140

Qy	1144	NOPDVRRPSPSPRCGLPLPAAPAGATTLERATLSPKNGGVVQVDFAPGAAVENPEYLTPQ	1200
Dd	1141	NQPEVRPOPPLTPREGPLPPVAPPAQATLERPTLLSPKNGGVVQVDFTPGAAVENPEYLTPR	1200
Qy	1201	CGAAPQHPHPAFSPAFDNLVYWDQDPBERGAPSTFGKTPTAENPEYLGLDVEV	1255
Dd	1201	CGSASQPH-PPALCGAFDNLVYWDQDPBERGSPPTFGKTPTAENPEYLGLDVEV	1254
RESULT 4			
EGFR_HUMAN	STANDARD:	PRT, 1210 AA.	
ID	EGFR_HUMAN	P00533; P06268; Q14225; Q9UMD7; Q9UM08; Q9UMG5; Q92795; Q00732;	
AC		O00688; Q9B282; Q14225; Q9GZC9; Q9GZC1; Q9H3C9;	
AC		01-NOV-1997 (Rel. 01, Created)	
DT		21-JUN-1986 (Rel. 01, Last sequence update)	
DT		15-JUN-2002 (Rel. 41, Last annotation update)	
DE		Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor	
DE		protein-tyrosine kinase ErbB-1).	
GN		EGFR OR ERBB1.	
OS		Homo sapiens (Human).	
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX		NCBI_TaxID=9606;	
RN		[1]	
RP		SEQUENCE FROM N.A. (ISOFORM 1).	
RX		MEDLINE=84219729; PubMed=6328312;	
RA		Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,	
RA		Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,	
RA		Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;	
RT		"Human epidermal growth factor receptor cDNA sequence and aberrant	
RT		expression of the amplified gene in A431 epidermoid carcinoma cells.;"	
RL		Nature 309:418-425 (1984).	
RN		[2]	
RP		SEQUENCE FROM N.A. (ISOFORM 2).	
RX		TISUE=Placenta;	
RC		MEDLINE=95382857; PubMed=7654368;	
RA		Ilekta J.V., Stark B.C., Scoccia B.;	
RT		"Possible role of variant RNA transcripts in the regulation of	
RT		epidermal growth factor receptor expression in human placenta.;"	
RL		Mol. Reprod. Dev. 41:149-156 (1995).	
RN		[3]	
RP		SEQUENCE FROM N.A. (ISOFORM 2).	
RC		TISUE=Placenta;	
RX		MEDLINE=97078686; PubMed=8918811;	
RA		Reiter J.L., Maible N.J.;	
RT		"A 1.8 kb alternative transcript from the human epidermal growth	
RT		factor receptor gene encodes a truncated form of the receptor.;"	
RL		Nucleic Acids Res. 24:4050-4056 (1996).	
RN		[4]	
RP		SEQUENCE FROM N.A. (ISOFORM 2).	
RC		TISUE=Placenta;	
RX		MEDLINE=97256547; PubMed=9103388;	
RA		Ilekta J.V., Gariti J., Niederberger C., Scoccia B.;	
RT		"Expression of a truncated epidermal growth factor receptor-like	
RT		protein (TRGFR) in ovarian cancer.;"	
RL		Gynecol. Oncol. 65:36-41 (1997).	
RN		[5]	
RP		SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).	
RC		TISUE=Placenta;	
RX		MEDLINE=21100872; PubMed=11161793;	
RA		Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,	
RA		Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,	
RA		Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,	
RA		Maible N.J.;	
RT		"Comparative genomic sequence analysis and isolation of human and	
RT		mouse alternative EGFR transcripts encoding truncated receptor	
RT		isoforms.;"	
RL		Genomics 71:1-20 (2001).	
RN		[6]	
RP		SEQUENCE OF 575-687 FROM N.A.	
RA		Reiter J.L., Thredgill D.W., Danielson A.J., Schehl C.M.,	
RA		Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,	

RA Maible N.J.:
RT "Human and mouse alternative EGFR transcripts encoding only the
RL extracellular domain of the receptor.";
RN Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
[7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326661;
RA Lin C.R., Chen W.S., Krueger M., Stolaraky L.S., Weber W.,
BA Evans R.M., Verna I.M., Gil G.N., Rosenfeld M.G.; gene
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RL cells.";
RL Science 224:843-848(1984).
[8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.D., Sullivan M., Wilson R.K., Ma D.P.,
RT Roe B.A., Merilino G.T., Pastan I.;
RL Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RN Nature 309:806-810(1984).
[9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
[10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=86217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
[11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1986448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
[12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merilino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
[13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
[14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "Atp-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
[15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Marzolis B.L., Lax I., Kris R., Dombalagian M., Honneger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlesinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HBR/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 264:10667-10671(1989).
[16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RA MEDLINE=66398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts."
 RL Growth Factors 13:121-132(1996).
 RL [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668.
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor."
 RL J. Biochem. 127:65-72(2000).
 RP PARTIAL SEQUENCE AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Oda M., Inagaki F., Iax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor."
 RL J. Biol. Chem. 273:11150-11157(1998).
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens."
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X00588: CA25240.1; -
 DR EMBL: U05089: AAB53063.1; -
 DR EMBL: U48722: AAC50802.1; -
 DR EMBL: U48723: AAC50804.1; -
 DR EMBL: U48724: AAC50796.1; -
 DR EMBL: U48725: AAC50797.1; -
 DR EMBL: U48725: AAC50798.1; -
 DR EMBL: U48725: AAC50798.1; -
 Query Match 45.3%; Score 3074; DB 1; Length 1210;
 Best Local Similarity 49.0%; Pred. No. 4.8e-159;
 Matches 620; Conservative 176; Mismatches 363; Indels 106; Gaps 21;

QY ASLSEFDIOIEVOGVLLIAHNOVROVPLORLIRIVGTOLFEEDNALVALDNGDPLNNTTP 128
 DB YDLSTLQIOTVAGVGLVLAALNVTVEIPLENDQIHGMNMYENSYALVALNSND----- 126
 QY VTGASPGGLRELQSLSTLEILKGVLLIQNPOLCYODTILMKDIFPKNNQALATLIDTNR 188
 DB ---ANKTLKELPMRNLOEILHGAVERFSNNPALCNVESIQMRDISSDPLSMNSMDFOQH 183
 QY SRACIPGSPMCKGSGSKWSESSDQSLRTVCAGGCA-RCKGRLPQYIKANSKFGITEL 247
 DB LGSQCKCPSPNCSGWAGBENCCKLTKIICAOQSCSRCKGSPSDCHNCAAGCTGP 243
 QY KHSQCLALCFHNSGICELHCPALVTYNTDFESMNPBGRFTFASCTVACPYYVLTSD 307
 DB RESDVLVCRKFRDAETCQDTCPRLMLYPTTYQMVNVEGKYSFATCYKKCPRYVYTD 303
 QY VGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLMEHLREAVTSAIOEFAC 367
 DB HGSQVRACGADSYEM-EDGVKCKCKGCPKRVKNGIGIEFKDLSINATINIGHFNK 362
 QY KKTIGSLAFLPESPDGPASTAPLQEPLOVETLEITGYLISAMPDLSPLDSVFON 427
 DB TSISGDHLIPLVAFRGDSFTHTPPLDPQELDLKTVKEITGFLLIQAMPENRTDLHAFEN 422
 QY LOVIRGRILHNGAVSLTQGLGISWLGRLSRLBGLSSALAIHNNHLCFVHVTPMDQFLR 487
 DB LEIRGRKQKQSGSLAVSLNITSLSRLSKEISDGVYIISGNKNLYCYANTINKKKILG 482
 QY NPHQALHTANRPDECEVGEGLACHQLCARGHGWPGPTQCNCQSLFRGQCEVECRVL 547
 DB TSGQKTKLISRGNSCQATQVCHALCSPEGCMGPERDVCSCNNSRGRCVCKNLL 542
 QY QGLREYVNAHCLPCHEPCOPNGSVTCFPEADQVCAAHYKDPFCVACRPSGVKPD 607
 DB EGPEREFVENSECICQCHPECLPQAMNITCTGRGPNICQCAHYIDGPHCVKTCFPAWGGE 602
 QY LSYPIWKFPEBEGACOPCPINCTHSCVDLDKGCAPAGRASPLTISYAVG---ILLV 664
 DB NNTL-VKRYADAGVCHLCHPNCYCTGCPGLBEGCTPNGPXP-R-SIATGAGALLLLLV 659
 QY VLVGVEGILIKRQOKIRKXTMYRRLQETELVEPLTPSGAMPNOQMRILKETELRYK 724
 DB VALDIG---LFMRRLHLYKXKTLRLLOERLVEPLTPSGAPNOALIRILKETEFKIK 716
 QY VLSGAGCTYVKGWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYSR 784
 DB VLSGAGCTYVKGWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYSR 776
 QY LIGICLSTVOLVQLMPYGLDLHVRNRRGLSODLLNMCQIAKMSLTEDRLVHR 844
 DB LIGICLSTVOLVQLMPYGLDLHVRNRRGLSODLLNMCQIAKMSLTEDRLVHR 836
 QY DLAAARNLVKSPNHVKTITDFGLARLIDIDETEVADGKVPKIMMALESIRRRTHOSD 904
 DB DLAAARNLVKTPGVKTTDFGLARLIDIDETEVADGKVPKIMMALESIRRRTHOSD 896
 QY VMSYGVTVVMSLMTFGAARYDGIIPAREIPDLIEKGERLPQPICTIDVYMIWVKCMI 964
 DB VMSYGVTVVMSLMTFGAARYDGIIPAREIPDLIEKGERLPQPICTIDVYMIWVKCMI 956
 QY CRPRFRELVEEFMSMARDPQRFVYIO-NEDLGPAASPLDSTYRSLBEDDDGDLVDAEY 1023
 DB SRPRFRELITEFSMAARDPQRYLVYIQDDEHRLPSPTDSNYRRLMOBEDDDVDVDAEY 1016
 QY LVPQGFCCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAGS 1083
 DB LVPQGFCCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAGS 1077
 QY DVFDGDLGMAAGLQSLPETHDPSLPORYSEDPVLPSET--DGYVAPLTCSPQPEYVN 1141
 DB DVFDGDLGMAAGLQSLPETHDPSLPORYSEDPVLPSET--DGYVAPLTCSPQPEYVN 1043
 QY N--NSTVACIDRNGLOSCPIKEDSFLORYSSDPGALTEDSIDDTFL-----FVPEYIN 1094
 DB N--NSTVACIDRNGLOSCPIKEDSFLORYSSDPGALTEDSIDDTFL-----FVPEYIN 1094


```

1142 QPDVRFPPSPREGSLPAAPAGATLERAKTSPKXGVVXQVFAFGAVENPEYL--TPQ 1200
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1095 Q-SVPRKPRGVSQNPVYHNPDPNP-----APSRDPHYD--PHSTAVGVPEYLVNTVQ 1143
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1201 GGAAPRPHPRPAFSPAFDNLVYWDQ-----DP-----PERGAPPSITFKGTPPAE 1244
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1144 -----PTCVNSTFDSPPAWAQKSHQSLNDPVDYODFPFKKAKPRGIFPKGS--TAE 119
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1245 NPEYL 1249
      | | | | |
1194 NAEYL 1198

```

```

      J 5      PRT; 1210 AA
      MOUSE
      -EGFR_MOUSE
      STANDARD;

```

001279; Createat) .
01-FEB-1996 (Rel. 33, Last sequence update)
01-FEB-1996 (Rel. 33, Last annotation update)
15-JUN-2002 (Rel. 41, Last annotation precursor (BC 2.7.1.112).
Epidermal growth factor receptor
EGFR.
Mus musculus (Mouse)
Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniomorphi; Muriidae; Murinae; Mus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
NCBI TaxId=10090;

[1] FROM N.A. GEORGE

SEQUENCE=1
STRAIN=BALB/c; TISSUE=Liver;
STRUT=81026370; Pubmed=1408137;
Givol D.;
factor receptor 2

MEBLINE=3500.
Avivi A., Skorecki K., Yayon A.
"Promoter region of the murine fibroblast growth factor-1 (bek/KGFR) gene.";
7.1957-1962(1992).

[2] Oncogene, 1990, 4, 111-115.

SEQUENCE=115302-
STRAIN=BALB/c, and CD-1; 1678348;
pubmed=7678348;
Dey S.K.;
factor gene is regulated

paria B.C., Das S.K., Andrews growth factor receptor of the epidermal delayed implantation."

"Expression of *...* during development of *...* in mouse blastocysts during development." *Natl. Acad. Sci. U.S.A.* 90:55-59 (1993)

PROC. NATL. ACADEMY OF SCIENCES
[3] FROM N.A.

SEQUENCE=1
STRAIN=BALB/c;
TISSUE=LIVER

RA Hibbs M.D.'s to the University of
Submitted (APR-1994)

RP	SEQUENCE FROM N.A.	TISSUE=liver;	0135255;	and N.G.	Earp H.S.
RN	[4]				

RC	STRAIN=20170986; PubMed=9122257
RX	MEDLINE=94170986; Phillips H.K., Qiu T.H., Copeland N.C., joint mutation in the EGF

RA
Jenkins N.A., Lee D.C.;
phenoty

RT "The most
receptor tyrosine kinase.
RT 8:399-413 (1994)
Genes Dev.

RL Genes
RN [5]
SEQUENCE OF 1-714 FROM N.A.

RP Sequence
Tissue=Brain;
RC Pubmed=2030916;
MentlInE=91232866; Glaxo D., Morse B.
Schlessinger J., Glaxo D., Morse B.
to study the ligand

Avivi A., Lax I., Ullrich
#Comparison of EGF receptor

binding site.";
oncogene 6:673-676 (1991).

RL [6] SEQUENCE OF 969-1117 FROM N.A.
RN
RT

RC STRAIN=C3H; serrero G.; EMBL/Genbank/DBD databases.
RC Finder D.P.; 1993) TO THE EMBL/Genbank/DBD databases.
RC Finder D.P.; 1993) TO THE EMBL/Genbank/DBD databases.

RL FUNCTION: THE EGF RECEPTOR MEDIATES AND HEPARIN-BINDING EGF, GROWTH FACTOR-RELATED PROTEINS (e.g., TRANSFORMING GROWTH FACTOR- α , AMPHIREGULIN, HEPARIN-BINDING EGF, GROWTH FACTOR-RELATED PROTEINS) AND OTHER LIGANDS TO INDUCE CELL GROWTH AND DIFFERENTIATION.

AND ALSO OF 1gF AND 1gM ANTIBODIES TO THE
VACCINIA VIRUS GROWTH FACTOR (BY SIMULTANEOUS
ADDITION OF ADP + A PROTEIN TYROSINE KINASE
INHIBITOR TO THE REACTION MIXTURE)

tyrosine phosphate. Type I membrane protein.
-1- SUBCELLULAR LOCATION: Type I receptor leads to
-1- MISCELLANEOUS: Binding of EGF to the receptor complex, cell DNA
dimerization, internalization of the EGF-receptor complex, cell
induction of the tyrosine kinase activity, stimulation of
proliferation and cell proliferation.
-1- NAME: EGF RECEPTOR FAMILY.

1- SIMILARITY: BELONGS TO THE...
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its use
 by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Use by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement>)
 entities requires a license agreement (See <http://www.isb-sib.ch/>)

or send an email to:

EMBL: X78987; CAA55587.1; -
EMBL: U03425; AAA17899.1; -
EMBL: X59688; CAA42219.1; -
EMBL: L06864; AAA53029.1; -
EMBL: Z12608; CAA78249.1; -
EMBL: P11362; 1FGK. -
HSSP: P11362; 1FGK. -
EMBL: P11362; 1FGK. -

MGD: MG1.952344; EGFR, L domain.
InterPro: IPR0000494, Euk_Dkinase.
InterPro: IPR000719, Euk_Dkinase.
InterPro: IPR002174, Furin-like.
InterPro: IPR001345, Tyr_kinase.
InterPro: IPR000695, Kinase; 1.
Pfam: PF00057, Furin-like; 1.
Pfam: PF00757, Furin-like; 2.
Pfam: PF01030, Recep_L domain; 1.
ProDom: PD0000001, Euk_kinase; 1.

DR F1A0261; FO; 3. 1.
DR SMART; SMO0219; TYRK; 1. KINASE ATP; 1.
DDR SMART; SMO0107; PROTEIN KINASE TYR; 1.
PROSTE; P500109; PROTEIN KINASE DOM; 1.
PROSTE; P550011; Receptor; Signal; Transferrase;
DR PROSTE; Glycoprotein; Receptor; phosphorylation; Repeat
Transmembrane; Glycoprotein; ATP-binding; phosphotylation; Repeat
KW Tyrosine-protein kinase; 24
KT SIGNAL 1
KT 25 1210
KT EXTRACELLULAR (POTENTIAL).

FT	CHAIN	25	647	POTENTIAL.
FT	DOMAIN	648	670	CYTOPLASMIC (POTENTIAL
FT	TRANSMEM	671	1210	APPROXIMATE.
FT	DOMAIN	75	300	APPROXIMATE.
FT	REPEAT	390	600	SER-RICH.
FT	REPEAT	1028	1071	PROTEIN KINASE
FT	DOMAIN	714	981	ATP (BY SIMILARITY).
FT	DOMAIN	720	728	ATP (BY SIMILARITY).
FT	NP BIND	747	747	SIMILARITY.

FT	ACT SITE	839	839	BY SIMILARITY.
FT	DISULFID	190	199	BY SIMILARITY.
FT	DISULFID	194	207	BY SIMILARITY.
FT	DISULFID	215	223	BY SIMILARITY.
FT	DISULFID	219	231	BY SIMILARITY.
FT	DISULFID	232	240	BY SIMILARITY.
FT	DISULFID	236	248	BY SIMILARITY.
FT	DISULFID	251	260	BY SIMILARITY.
FT	DISULFID		291	BY SIMILARITY.

PT	DISULFID	264	307	BY SIMILARITY.
PT	DISULFID	285	326	BY SIMILARITY.
PT	DISULFID	311	333	BY SIMILARITY.
PT	DISULFID	329	515	BY SIMILARITY.
PT	DISULFID	506	523	BY SIMILARITY.
PT	DISULFID	510	535	BY SIMILARITY.
PT	DISULFID	526	535	BY SIMILARITY.
PT	DISULFID	539	555	BY SIMILARITY.
PT	DISULFID	558	571	BY SIMILARITY.
PT	DISULFID	579		

FT	DISULFID	562	591	BY SIMILARITY.
FT	DISULFID	582	617	BY SIMILARITY.
FT	DISULFID	595	628	BY SIMILARITY.
FT	DISULFID	620	636	BY SIMILARITY. (BY PKC) (BY SIMILARITY)
FT	DISULFID	624	680	PHOSPHORYLATION (BY SIMILARITY).
FT	MOD_RES	680	1092	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1092		

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
 (BY SIMILARITY).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 19 19 C -> S (IN REF. 2).
 FT CONFLICT 539 539 C -> W (IN REF. 5).
 FT CONFLICT 991 991 L -> F (IN REF. 4).
 FT CONFLICT 991 991 HP -> DR (IN REF. 6).
 SQ SEQUENCE 1210 1117 AA; 134853 MM; 690E20D46DF202F5 CRC64;

Query Match 45.0%; Score 3053; DB 1; Length 1210;
 Best Local Similarity 49.0%; Pred. No. 6.5e-158;
 Matches 623; Conservative 169; Mismatches 370; Indels 110; Gaps 23;

QY 11 LLLALLPFGAA--STOVCTGTMKLRLPASPTHLMRLHYQCQVQVQNLLELTYPN 68
 DB 14 LLLALLPFGAA--STOVCTGTMKLRLPASPTHLMRLHYQCQVQVQNLLELTYPN 73
 QY 69 ASLSFLDIOGVQVLIANNOVROVLOLRIVRGVQLPFDNYALAVLNDGDLNNTTP 128
 DB 74 YDLSFLTIQEVAGVLIANTVERIPLNLIIRGNALYENTYALALISL----- 124
 QY 129 VTGASPGELRELQSLTEILKGGVLIQNPOLCYOPTILMKOI-----FHKNNOLATLI 184
 DB 125 -YGTNRGLRELPRNIOELIGAVRFSNNPILCMNTIQRDLYQNVFMSMDL---- 180
 QY 185 DYNRSRACHPCSPWCKSGRCSSSESDCQSLRTVCAAGCA-RCKGRLPOYIKANSKFIG 243
 DB 181 -QSHPSSCPKDCSPWCKSGRCSSSESDCQSLRTVCAAGCA-RCKGRLPOYIKANSKFIG 239
 QY 244 ITLKHSDCLACILHFNISGICELHCPALVNTDTFFESMPREERTFGASCYACPRNY 303
 DB 240 CTGPRESDCVLCQKFODEATKDCPPMLVNTDTFFESMPREERTFGASCYACPRNY 299
 QY 304 LSTDVGSCTLVCPFLHNOEVTAEADTORCEKSKPCARCYGLGMEHREVAVTSANIOE 363
 DB 300 VVTHDGSCTVACGPDYEV--EEDGIRKCKCDEGCRKVCNGIGIGERKDTLSIATNIKH 358
 QY 364 PACCKKIFGSLAPLESFDDPASNTAPLOEQLQVETLEITGYLYISAMPDLPDLS 423
 DB 359 FKYCTAISGDLHLPVAFKGDSEFTPTPLDRELEIKTVKEITGFPILLQAMPDMDLH 418
 QY 424 VPONLOVIRIRILHNGVSLTLOGLGSMGLRSLRELGSGLAIHNTHLCFVHTYPMW 483
 DB 419 AFENLEIRIRTKOHQGFSLAVVGLNTSLGLRSLKISOGDVIISGRNLVYANTINWK 478
 QY 484 QLEFRNPQALLHTANRDECEVCEGLACHOLCARGHGCMGPQPTCVNCSQFLRGCEVEE 543
 DB 479 KLFCTNRQTKKIMNRAREKCKCAVNHVNCPLCSGEGMGEPKRCVSCQNVSRKECEVEK 538
 QY 544 CRVLQGLPREYVNAHCLPCHPECPONGSVTCGPEADOCVACAHYKDPFCVAPCPSG 603
 DB 539 CNLIEGEPREFEVNSECTICQHPRECLPQAMNITCTGRBDNCIOCAHYIDGHCYKCTPAG 598
 QY 604 VKRDLVYMPKPRDEGACQPCPINTHSCVLDLDBGCAPEAOASLTSIVSAVVGILL 663
 DB 599 IMENNTL-VWKYADANNVCHLCNANTYGCAGBGLQGEVWSPGPRIPATIGVGLL 657
 QY 664 VVVLGVVFGI-LIKRROOKIRKTYMRLLOTELVEPLTSGAMPNOAKRIKETEPLK 722
 DB 658 FIYV-VALGIGLPMRRHIVRKTKRLRLDLRELEVEPLTSGEAPNPAHRIKLEITEFKK 716
 QY 723 VKVLGSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPRANKELIDEAYVMGVGSPYV 782

DU 717 IKVLGSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPRANKELIDEAYVVASVDNPHV 776
 QY 783 SRLGICLSTVOLVTOUMPYGCLLDHYRENRGRGSDGLNCKMCIKMSYLEDVLY 842
 DB 777 CRLGICLSTVOLVTOUMPYGCLLDHYRENRGRGSDGLNCKMCIKMSYLEDVLY 836
 QY 843 HRDLAARNVLVKSPNHVITDPLGLARLLDIETETENHAGGVKPIKMALESILRRFTHQ 902
 DB 837 HRDLAARNVLVKSPNHVITDPLGLARLLDIETETENHAGGVKPIKMALESILRRFTHQ 896
 QY 903 SDVMSYGVTVWELMTFGAKPYDGI-PAREIPDLEKGERLPQPICTIDVMIWVCMID 962
 DB 897 SDVMSYGVTVWELMTFGSKPYDGI-PASDISILEKGERLPQPICTIDVMIWVCMID 956
 QY 963 SECPRRRELVSERKARQDRVVIO-NDLGPASPLDSTFRSLLEDMDGLVDAE 1021
 DB 957 ADSRPKRELLEFSKARQDRVLVIOGDRLMPLSPPTDSNFYALMDEEMEDVVDAD 1016
 QY 1022 EYLVPOQGFPCPDAPGAGVHHRHSSSTRSGGDLTLGLEPSEEPAPSPPLAPSGA 1081
 DB 1017 EYLVPOQGF-----NSPST-----SRPLSSLSA 1042
 QY 1082 GSDVFDLGLGAKGQLSLPTHDPSPLOQYSEDPVPLPSET--DGVAPLTCSPOEY 1139
 DB 1043 TSN-----NSTVACINRNGSCRVKEDAFLOQYSDPTGAVTEDNIDDAFL-----PVPEY 1092
 QY 1140 VNQPDVAPPPSPREGLPAPAPAGATLERAKTISPCKNGVAVKQVAFAGAVENDEYL-T 1198
 DB 1093 VNQ-SVPRKPAQSVQNPVYHNPQLHP-----APGRDLHYQN--PHSNVGNDEYNT 1141
 QY 1199 POGGAQPPPPAPFSPAFDNLVYWDQ-----PP-----PENGAPSPFFKGPPT 1242
 DB 1142 AQ-----PTLSSGFSFSLMILQKSHQMSLNDNDYQDFFPKETKNGIFKG-PT 1191
 QY 1243 AENPEYGLDVP 1254
 DB 1192 AENPEYLRVAP 1203

RESULT 6
 ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
 AC QLS03;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=838326;
 RA Plozman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 RA Foy L., Neubauer M.G., Shoyab M.;
 RT "ligand-specific activation of HER4/p180erbB4, a fourth member of the
 RT epidermal growth factor receptor family";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester";
 RL J. Biol. Chem. 272:26761-26768(1997).


```

Db      536 CNLYDGEFREFNSICVECDPCEKMEGDLTCHGPGPDNCTKSHFKDGNVEKCPD 595
Qy      603 GVKPDLSPYIMKPEDEGACOPCPINCTHSCVDLDDKGC-----PAEGRASPL 651
Db      596 GLOQANF--IFKXADPRBECHPCHPNCTOGCNPSTSDCIYPTWGTSTLPQAR--TPL 652
Qy      652 TSIVSAVV-GILLVVLGVFGLIKRRQOKIRKYMRLLQETLVEPLTPSGAMPNQA 710
Db      653 --IAAGVIGGLFLLVIVGLTFAVVVRRRSIK-KKRALRRFL-ETELVEPLTPSGAPNQA 708
Qy      711 QMRLLKTELKRVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSKANKEILDE 770
Db      709 QURLKTELKRVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSKANKEILDE 768
Qy      771 AYYAGVGSPPVSRLLGICLTSTVOLVQLMPYGLLDHVENRRLSGODLLMCMQIA 830
Db      769 ALIAMSADHRLVLLGVCLSPITQLVQLMPHGLLEYVHNKDNIGSOLLIMCVQIA 828
Qy      831 KMSYLEDVRLVHRDLAARNVLYKSPNHVKITDPLGLARLDIDETEVHADGKVPKIMA 890
Db      829 KGMVYLEERLVRHDLAARNVLYKSPNHVKITDPLGLARLDIDETEVHADGKVPKIMA 888
Qy      891 LESTLRBFTHOSVMSGYVTWELMTFGAKPYDGIAPARELPDLLEKGERLPQPICTID 950
Db      889 LECIHYKFTHQSVMSGYVTWELMTFGAKPYDGIAPARELPDLLEKGERLPQPICTID 948
Qy      951 VYIMVAVCMWIDSECRPFRELVESEFMRADPQRFVVIQONED-LGPASPLDSTFYRSL 1009
Db      949 VYIMVAVCMWIDSECRPFRELVESEFMRADPQRFVVIQONED-LGPASPLDSTFYRSL 1008
Qy      1010 EDDMDGLVDAEYLVPOQGFCCDPAPAGAGVHHRRSSSTSGGDLTLGLEPSEE 1069
Db      1009 DEEDLEMDMAEEYLVLP-QAFENIPPP-----IYTSRARIIDSNRS-----EIGHSPRAY 1056
Qy      1070 APRS-----PLAP-SEGAGSVDFDGLDMGAKGLQS 1100
Db      1057 TPNSGNOFVYRROGFAAEQGVSVYRAFTSTIPEAPVAGATAEIPDSCCNGLTRKVA 1116
Qy      1101 LPTHDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQPEYVQNPVRQPPSPR 1153
Db      1117 PHVEDSESTQRYSDPTVYFAERSPRGELDEGYMTPRMDPKQEYLVNPV----- 1167
Qy      1154 EGPLAPAPAGATLERAKTSLSPKNGGVKQVFAFGAVENPEYVLPQGAAPQPHPPA- 1212
Db      1168 ENPVSRR-----KNGDLQ-----ALDNEYNHNASNG-----PPKAE 1199
Qy      1213 -----FSPAFDNLYMDQPPERGA--PPSTF 1237
Db      1200 DEYVNEPLYLNTFANTLGKAEYLNKNTLSMPEKAKKAFDNDYWNHSLPPRSTLQHPYL 1259
Qy      1238 KGTPT-----AENPEYL 1249
Db      1260 QEVSTKTFYKQNGRIPIVAENPEYL 1285

```

```

RESULT 7
ERRB4_RAT
AC 062956; 0922N7; STANDARD; PRT; 1308 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERBB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnet P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997)
RN
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NR-
CC 2, NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as there is no content in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF041838; AAD08899.1; -.
DR EMBL; U52531; AAC53051.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR0000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TYKIC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR TRANSFAC; PS00109; PROTEIN_KINASE_TYR; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT CHAIN 1 25 POTENTIAL.
FT FT 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE
FT NP BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-)
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-)
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-)
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-)
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .)
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .)
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .)
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .)
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .)
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .)
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .)
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .)
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .)
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .)
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944B096A08B41 CRC64;

Query Match 42.6%; Score 2890; DB 1; Length 1308;
Best Local Similarity 44.8%; Pred. No. 4,7e-145;
Matches 603; Conservative 189; Mismatches 395; Indels 160; Gaps 28;
```

```
QY 357 TSANIOEFAGCKKIFGSLAFLPESFPGDPASNTAPLQEOLOVFETLEITGYLISAMP 416
DB 349 DSSNIDKINCKTKINGNIIFLVGTGHGDPYNAIDIDBEKLVFTVAREITGFLNIQTWP 408
QY 417 DSLPDLVFNQLOVIRGRIHLNGAVSLTLQGLISWGLRLSRLGSLGLAIHNHTLICE 476
DB 409 PNMTDFSVFSLVLTIGRVLVSGLSLILIKQGGITSLQFOSILKEISAGNIYITDNSNLCY 468
QY 477 VHTVPMQDLFNNPQALHTANRPDECEVGLACHOCARGHCHGPRPTOCVNCISOFLR 536
DB 469 YHTIMVTLTFSTVQRIYIRDNRRARENCTAGAGMVCNHLCSNDGCMGPDPDCLSCRRFSR 528
QY 537 GOECVEECRVYQGLPREYVNAHCLPCHPECCP-ONGSVTCGCPADOCVACAHKXDPF 595
DB 529 GKICIESNVLGDEREPENGSIYCECCSQCEKMEGDLITGCHGPDPDCTKSHRKDGN 588
QY 596 CVARCPGKVDLSYMPIKPEPDEBGACQPCPINTHSCVLDLDDKGC-----PA 644
DB 589 CVEKCPDVLQANSE--IFKYADQRECHPCHPNCTGCGNCPSTSHDCIYYPMTGHSITLPQ 646
QY 645 EORASPLTISYSAV-GLLVVVLGVVFGILIKRQOKIKRKYTMRLLOETLVEPLTPS 703
DB 647 HAR-TPL--IAAGVIGLFLIVMALTPAVVYRRKSIK-KKRALRFL-ETELVEPLTPS 701
QY 704 GAMPNQAMRIKETELRKRVKVLGSGAFGYVKGIMIPDGENVKIPVAIKVLRNTSPKA 763
DB 702 GTAPNQADRLIKETELRKRVKVLGSGAFGYVKGIMIPDGENVKIPVAIKVLRNTSPKA 761
QY 764 NKEILDEAYVAVGSPYVSRILGICLTSTVQVLTOLMPYGCGLDHYENRGRLOSL 823
DB 762 NYEFDEALIMASVDHPHLVGLVCLSPITQLVQLMPHGCGLLEYVHEHKNIGSOQLL 821
QY 824 NMVQOIAAGMSYLEDVRLVHRDLAARNLVKSPNNVKTITDGLALLDIDETEVYADGK 883
DB 822 NMVQOIAAGMSYLEDVRLVHRDLAARNLVKSPNNVKTITDGLALLDIDETEVYADGK 881
QY 884 VPIKMALESILRRFTHQSDVWSYGVTVMELTGAKPYDGI PRAEIPDLLEKGERLPQ 943
DB 882 MPKIMALECHYKRFTHQSDVWSYGVTVMELTGAKPYDGI PRAEIPDLLEKGERLPQ 941
QY 944 PPICTIDVYIMVCMIDSECRPFRELVSFESMABDPORFVYIQNED-LGPASPLDS 1002
DB 942 PPICTIDVYIMVCMIDADSRPFKELAEFSMABDPORFVYIQNED-LGPASPLDS 1001
QY 1003 TFRSILEDDMGDLVDAEVLVPOGCFPCDP-----ARCA 1039
DB 1002 KPFQNLDEEDLEDDMDAEVLV-QAENIPPIYTSRTIDSNRSEIGHSPPAYTPMS 1060
QY 1040 GAMVYHRRSSSTRSGGDLTLGLEPSEEBAPRSPLAPSEGAQVDFQDLCMGAKALQ 1099
DB 1061 GSGPVLQDGGFATQGG---MEMPTATITSTIPBAVVA--QATAMFDDSCNGTLRKPV 1115
QY 1100 SLPTHDPSPLORYSEDPVPLPS-----ETDGVAVRLTCSPOPEYVNOQDPVRQPPSP 1152
DB 1116 VPHVQEDSSTQYSADPTVPAPERNRPAELDEBEGVTBMHDKPKQEVLNPLYE----- 1167
QY 1153 REGPLAPRPAAGATILERAKTLSPGKNGVYKQVFAFGAVENPEYLTPGGAAPQHPPPA 1212
DB 1168 -ENPVSRR-----KNQDLQ-----ALDNPETHSASSG-----PPKA 1198
QY 1213 -----FSPAEDNLVYWDQDPPERGA--PPST 1236
DB 1199 EDEYVNEPLVNTFTNALGNAEYMKNSLSPBEKAKKAFDNPDPVYNNHSLPFRSTLQHPDY 1258
QY 1237 FKGTPT-----AENPEYL 1249
DB 1259 LOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 8
XMRK_X1PWA STANDARD; PRT; 1167 AA.
AC P13388;
```

DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
OS XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxId=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Mutchrodt J., Adam D., Malitschek B., Mauelel W., Raulf F.,
RA Telling A., Robertson S.M., Scharlt M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Scharlt M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16891; CA34770.2; -
DR PIR; S06142; S06142.
DR HSSP; P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recept_Ldomain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TY-TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167
FT FT 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
FT FT 26 1167 KINASE.
FT FT 26 1167 EXTRACELLULAR (POTENTIAL).
FT FT 26 1167 TRANSMEM 643 665 POTENTIAL.
FT FT 26 1167 DOMAIN 666 665 CYTOPLASMIC (POTENTIAL).
FT FT 26 1167 DOMAIN 710 977 PROTEIN KINASE.
FT FT 26 1167 NP BIND 716 724 ATP (BY SIMILARITY).
FT FT 26 1167 BINDING 743 743 ATP (BY SIMILARITY).
FT FT 26 1167 ACT_SITE 835 835 BY SIMILARITY.
FT FT 26 1167 DISULFID 195 204 BY SIMILARITY.
FT FT 26 1167 DISULFID 220 228 BY SIMILARITY.
FT FT 26 1167 DISULFID 224 236 BY SIMILARITY.

FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
Query Match 38.7%; Score 2625.5; DB 1; Length 1167;
Best Local Similarity 44.8%; Pred. No. 8.56-135;
Matches 568; Conservative 162; Mismatches 400; Indels 137; Gaps 27;
OY 4 AALCRMGILLALALPPGAAS- - - - - QVCTGDMKRLRASPETHLDMLRHLCQCGVOGON 59
DB 8 AALIQ--LLVLVLSIRCCSTDPDRKVCQGSINQTM--LDNHLVKMKMVGSCNVLVEN 62
OY 60 LELTYLPTNASTFLDIOIEVOGYVLLAHNOVROVPLQRLRIVRGTOLEFDNALVALVDN 119
DB 63 LEITYQENDLSLQIEVGCVLLTAMNEVSTPLVNLRLINGONLIEGNFTLLWSN 122
OY 120 GDPPLNTTPVTGASPGGLRELQRLSLTEILKGVLYIQRPOLCYODTILMKDIFHKNOQL 179
DB 123 YQK-NPSSP--DYVQVGLKQLOLSNLTETLISGGVKSHPNPLLCVETITMMDDIVDKTSNP 179
OY 180 ALTLIDNRSRACHPCSPMCKGSRCKWSSSDCSQSLRTVAGGC-ARCKPLQYITAN 238
DB 180 TTNALIPRAFERQCKCHGCVNGSCMAPRGHCQKFTKLCAECCNRRCRGPKPDDCNE 239
OY 239 SKFIGITELKHSDELACLFHNSGICELHCPALVTYNTDFESMNPREGRTFGASCYTA 298
DB 240 HCGAGCTGPRATDCLCRDPRNDSTCKDTCPRPKIYDISHQVVDNPKIKTFGAACVKE 299
OY 299 CPNYLSTDVSGCTLVCPRLHNOEVTADGTORCEKSPCARVCYGLGEMHLEVRVATVS 358
DB 300 CPNSYVYTE-GACVRSACAGLEVD-BNGKRSCKPCDGVCPKVDGIGISLSTIAVNS 357
OY 359 ANIQEPAGCKKIGSLAFESPFGDPASTAPLQPPQLOVFELEITGLYLSAMPDS 418
DB 358 TNIRSFNCKTKINGDILLNPNSEGDPRHYKIGTMDPRLNLTIVKXETGLVLTMMWEN 417
OY 419 LPDLSPONTQVIRGRILHNGAYS-LTLQGLISWLGSLRSLGSGALTLHNHTHCFV 477
DB 418 MTSLSVQNLLEIRGRTRTFRGRFSFVVQVRHLOMLRLSKLKEVASAGVILAKTLQRLYA 477
OY 478 HTVPMQDLFPNPHQALHTANRPBDECVGGLACHQICARGHCMGPPPTQCVCNSQFLRG 537
DB 478 NITNMRLFRSEDSIEYDART-----ENQTCNNCSBDGCGPPPTMVCSCLDHVRG 530
OY 538 QECVEBRYVQGLPREVYNARHCLPCHREGCPONGSTCGPREADQCAAHYDPPFCV 597
DB 531 GRCVASCNLQGEPRERQVDGRVCQHQECLVGTDTSLCTCGPREPANSKSAHFQDGQCI 590
OY 598 ARCPGKVPDLSTYMPIMKFPDEBACORPCINCTHSCVDLDDKGCAPRGRASPLTISVSA 657

```

Db      591 PRCHGILGDDDLT-IMKYADKMGCCQPCCHQNCCTGCGGPGLSGGRD-IYSHSLAAGL 648
Oy      658 VVGLLVVLCVFGILIKRROOKRKTYMRLLLOETLVEPLTPSGAMPNOQRIKE 717
Db      649 VSGELITVIVALLIVLLRRRRIRK-RKRTIRCLLOEKELVEPLTPSGAPNOQAFRIK 707
Oy      718 TELRKVKVLSGAGCTYKGIWIPGGENVKI PVATKVLRENSPANNEIIDEAVMGV 777
Db      708 TEFKDRVLGSGAGCTYKGLNPNGENIRIPVATKVLREATSPKVNQDEVAVMASV 767
Oy      778 GSPVYSRLIGICTSTVOLVTLMPYGLLDHVRNRRGLSGQDILAMCMQIAKMSYLE 837
Db      768 DHPHYCRLLGICTSAVOLVTLMPYGLLDVVRHQRIICQGMILNMCVOIAKAMNLE 827
Oy      838 DVLVHRDLARNVLVKSPNHYKITDGLARLLIDETRYHADGGRVPIKMALESIIIR 897
Db      828 ERHLVHRDLARNVLLKNPNHYKITDFGLSKLLTADKEKEYQADGGKVDIKMALESIIQW 887
Oy      898 RFTHSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRPLPOPICTIDVYMIWK 957
Db      888 TYHOSDWSYGVTVWELMTFGSKPYDGI PAEILASVLENGERLPPICTILEVMIILK 947
Oy      958 CWMIDSECRPFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGL 1017
Db      948 CWMIDPSRPRFRELVSFSGMARDPSRYLVIQ--NLPSISDRLLFSRLSSD--DV 1002
Oy      1018 VDAEYVLVPOQGFPCPPDPAAGGVHRRHSSSTRSGGGLTGLPSEEBEAPSPILAP 1077
Db      1003 VDAEYVLVLYKRI-----NRQSS-----EPCIP 1025
Oy      1078 SEGAGSDVDDGLGMAKAGLSLPTHDSPLOQRYSEDPV-PLPSETDGYVAPLTCSPQ 1136
Db      1026 PTGH-----PVRENSITLRKISPTQNALSKDDGH----- 1056
Oy      1137 PEYVNPDPVRPQ-----PSPRE-----GPLP-AAPAGATLERAKTLSPKKNVYKD 1183
Db      1057 -EYVNPQSESTSRSLDIYNPYEDLTQMGVSLSGQEAETNFSRPEVLTNNQSL--- 1112
Oy      1184 VFAAGAVENPEYLTPOGGAAPQHPHPAFSPAFNLYYMDODPPERGAPSTFGTPTA 1243
Db      1113 PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTALTGNGMFLPA 1150
Oy      1244 ENPEYIG 1250
Db      1151 ENLEYLG 1157

RESULT 9
ERB3_HUMAN
ID_ERB3_HUMAN STANDARD: PRT: 1342 AA.
AC P21850.
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Klaus M.H., Issing W., Miki T., Pinescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RP [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

```

```

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RT Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine kinase + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC - ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC - DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M29366; AAA35790.1; -.
DR EMBL, M34309; AAA35979.1; -.
DR EMBL, S61953; AAB26935.1; -.
DR PIR, A36223; A36223.
DR HSSP, P11362; 1FGK.
DR Genew, HGNC:3431; ERBB3.
DR MIM, 190151; -.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam, PF00069; Pkinase; 1.
DR Pfam, PF00757; Furin-like; 1.
DR Pfam, PF01030; Recep_L domain; 2.
DR Prodom, PD000001; Euk_kinase; 1.
DR SMART, SMO0261; FU_3.
DR SMART, SMO0219; TyKc; 1.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP. FALSE NEG.
DR PROSITE, PS00109; PROTEIN_KINASE_TYR. FALSE NEG.
DR PROSITE, PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1..19
FT CHAIN 20..1342
FT DOMAIN 20..643
FT TRANSMEM 644..664
FT DOMAIN 665..1342
FT DOMAIN 709..966
FT NP_BIND 715..723
FT BINDING 742..742
FT ACT_SITE 834..834
FT DISULFID 186..194
FT DISULFID 190..202
FT POTENTIAL.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE.
FT ATP (BY SIMILARITY).
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.

```



```
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASAPLIC 141 183 ELISGVYIEKDKLCHMDITDMDRIYVDRDEIYKONGR
SC -> GQPMVPSGLTQPMODWTLDLDDPDLTLTSSSK
VPTLAIV (IN SHORT ISOFORM).
FT VASAPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E766CA374BD CRC64;
```

Query Match 34.7%; Score 2354.5; DB 1; Length 1342;
Best Local Similarity 40.1%; Pred. No. 4.6e-120;
Matches 526; Conservative 190; Mismatches 466; Indels 129; Gaps 32;

```
10 GLLALLRPGAA--STGYCTGDMKRLPAPETHLDMLRHLXGCGVYQGNLEIYPT 67
11 GLSLSLARGSEVNSQAVCPGTLNGLSTYDAENGYOTLYKLYERCEVVMGNLEIYVLTGH 70
68 NASISFLDIQEVGYVLIANOVQVPLQRLRIYRGTLFEDNVALLAVLNDPILNNT 127
71 NADISFLQIMREVGYVLIANNEFTLPLPILRVYRGTYVDGKFAITVM-----LNTNT 125
128 PVTGASPGGLREIQLRLSTLILKGGVLIQRNPOLCYODTILMKDIFHKMQLATLTLDTN 187
126 ----NSSHALRQLRLTOLTEILSGVYIEKNDKLCMDITDMDRIYDRD---AEIVKXD 178
188 RSPACHSPKCKSRCKGSESDCOSLITRVACGC-ARCKGRLPYQIYKANSKFIGITE 246
179 NGRSCPCHVEYCKG-KCMGPSESDCQTLTKTICAPQCGHCFGPNPNCCHDEACAGCGSG 237
247 LKISDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYFGASCVTACPYNTLT 306
238 PPDIDCACHFNHNSGACVPRCPPLVYNTKLTQLEPHNPKTYQYGVCAVSCPHNFV-V 296
307 DVGSGTLVCPLNHNEVTAEDGTORCEKSKPCARVCGYLGEMHLEVAVATSANIOEFAG 366
297 DQTSVCAVACPDDKMEVD-KNGLKMCCEPCGGLCPKACCEGTGSG--SRPTVDSNIDGFVN 353
367 CKKIFGSLAPFESFDDPASNTAPLOPELOVETLEETIETGYIYISAMPISLDLSVFO 426
354 CTYILGNLDFLITGLNDPWHKIPALDPEKLVNFRYAEITGYLNTISWPFHMFNFSVFS 413
427 NLQVIRGRIILHNGAYS-LTLQGLGTSWLGRLSRLRELSGLALIHNTHTLCPVHTVPMDL 485
414 NLTTIGRSLYNGFSLIMKNLNVSLGFRSLKEISAGRIYISANQOLCTHHSNMTKV 473
```

```
QY 486 FENPQALLHTA-NRPEDECVGEGLACHQLCARGHGMGPGPTQVNCOSFLRGQECVEEC 544
DB 474 LRGPTEERLDIGNRPRRCVABEYKVCDFLSSGGCGMGPQGLSCRYNSKGVCTHIC 533
QY 545 RVLQGLPREYVNAHRLCPHPECOPONGSVTCFGRBADQVCAAHYKDPFVCARCPSGV 604
DB 534 NFLNEPPEFAHEACFCSCHPCEQMEGTATCGSGSDTCAOAHFRDQPHCVSCPHGV 593
QY 605 KPDLSTYMIKRPDEGACORPCINCTHSQVLDLCKGCRABQRA----SPLTSIVSAVVG 660
DB 594 LG--AKGPIKYPRDQNBRCRPHENCTQCKGPELDQCGQTLVLIGKHTLMTALVING 651
QY 661 ILLVVLGVFGLIKRRQOKIR-KYTMRLIQLTEILVEPLTPSGAMPQAOIRILKETE 719
DB 652 --LVVI FMMLGSTFLYMGGRITQNGRARRRYLERESTIEPLDP--EKANKYLARIKETE 708
QY 720 LRKVKVLGSGAAGTYKGIWIPDGENVKI PVAIKYLRENTSPEKANKIIDEAYVMAGVS 779
DB 709 LRKLVLGSGVFGVTHKGVMIPEGESIKIPVCIKYIEDKSGHQSFOAVTDHMLAIGLDH 768
QY 780 PVSRLIGICLITSTVQLVQLMPVGCCLDHYRENGRGLSODLLMWMQIAGKMYLEEV 839
DB 769 AHIVRLGLCPRESSIQTLVQYPLGLSLDHYRQHFGALGPQLLMMVGQIAGMYLLEH 828
QY 840 RLVRDLAARVLYKSPHNVKITDPGLARLIDIDETERYHADGKYPKIMMALESILRRF 899
DB 829 GMVHRNLARVLYKSPQOVQVADRGVADLPRDDKQLLYSAKPIKIMMALESIHFGKY 888
QY 900 THQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPRPICTIDVYMIWVKM 959
DB 889 THQSDVMSYGVTTWELMTFGAEPYAGRLAEVPDLLEKGERLAQOICTIDVYMWVKM 948
QY 960 MIDSCRRPRELVSFESRMARDPQRFVIONEDGPA---SPLDSTFRSLLEDDMD 1016
DB 949 MIDENIRTFEELANEFPMARDPRYLVIKRES-GFGIAPRPEHGLTNKKLEVELEPR 1007
QY 1017 LVDAEYLVLPQGFPCPDPAAGAGMVAHRRSSSTRSGGDLTGLER-SEEAAPRSL 1075
DB 1008 ELDDLDLEAED-----NATTLTGLSLSPVGLTNRRPSQSLL 1048
QY 1076 APSEGAGSDVPDGDGMGAAGLOSPLTH-PSPLQRYSEDPVLP-----SETDGV 1128
DB 1049 SPSSGY-MPMNQGNIGESQSAVSGSSERCPRVSLH-----PMRGLCLASSESEGV 1101
QY 1129 A-----PLTSGPOPE---YVNPDVAPQPPSPREGP-----L 1157
DB 1102 TGSEAEIQEKVSKCRSRSRSPRPRGSAVHQSRLTLTPVPLSPGLEEEDVNGVYM 1161
QY 1158 PAARPAGATLEAKTLSP-GKNGVY-----KDVAFGAVENPEYLTPOGGAARQPPRP 1210
DB 1162 PDTHLKGTSSRREGTLSSVGLSVLGTBEEDB-----BEYETMRRRRHSP-PHPP 1212
QY 1211 PAFSPAEDNLYWD-----ODPERGAPPTSPFGKPTPAENPEYL 1249
DB 1213 RPSLSLELGEYMDVGSLSLSLSTGSCPLHPVPIMPTAGTTPDEDEVEM 1263
```

RESULT 10
ERB3_RAT
ID ERB3_RAT STANDARD; PRT; 1339 AA.
AC 062799; 062955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (C-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CN NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein."
RL Gene 165:279-284(1995).
RN (2)
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (Dec-2001) to the EMBL/Genbank/DBD databases.
RN (3)
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuoregulin and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL collaboration -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-db.ch/announce/>
or send an email to license@isb-db.ch).
CC -----
DR EMBL: U29339; AAC28498.2; -
DR EMBL: U52530; AAC53050.1; -
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001345; Tyr_kinase.
DR Pfam: PF00069; Kinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR Prodom: PD000001; Euk_kinase; 1.
DR SMART: SM00261; FU_5_kinase; 1.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
FT Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 662 POTENTIAL.
FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 259 CYS-RICH.
FT DOMAIN 707 964 PROTEIN KINASE.
FT NP_BIND 713 721 ATP (BY SIMILARITY).
FT BINDING 740 740 ATP (BY SIMILARITY).
FT ACT_SITE 832 832 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1028 1028 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AASF2402BBFD1E CRC64;
Query Match 33.6%; Score 2283.5; DB 1; Length 1339;
Best Local Similarity 40.4%; Pred. No. 3,2e-116;
Matches 518; Conservative 170; Mismatches 440; Indels 155; Gaps 34;
QY 3 LAALCRKGLLALPRGAA---STVCTGTDMKRLRPSPEPHLMDRLHYGCGVQVGN 59
DB 7 LQVLC---FLTSLARGSEMNSQAVCPGTLNGLSVTGDADQYQTLVLYKEKCEVGMGN 62
QY 60 LETLTPTNASHSFLODIOEVQGYLIHANOVRQVPLRLRIVRGTOLEFEDNYALAVLDN 119
DB 63 LEIVTIGHNADSLFQIMREVTGYLVAMNESVPLRLRLKRVKGTQYVDGFAIPVM-- 120
QY 120 GDPLNNTTPVTGASFGRLREIQLRSLEILKGVLIQRNPOLCYODTILMKDIFPKNNOL 179
DB 121 ---LNYNT-----NSHALRQLKFTQTEILSGVYIENKDKLCHMDTIDMRDIVVR-- 170
QY 180 ALTLIDTRSRACHCSPMKCSRCWGSSESCGLSTRVCAAGC-ARCKGLPOLYIKAN 238
DB 171 GAELIVKXNNGANCPCPCHEVCKG-RCMGPGPPDCCQILITIKICAPQCNCRGCFGPNPQCCHD 229
QY 239 SKFIGITELKSDCLACLPFNHSGICELHCPALVYNTDTFESMNPBEGRYFGASCYTA 298
DB 230 ECAGSGSPQDIDTCACRPFNDSGCVRPCPEPLVYNTKLTFLBEPHPTKYTGVCVVAS 289
QY 299 CPYNTLSTDVSGCTLVCPLEHNOEVTABDGTORCEKSKRCARVCYGLGMEHLREVAVTS 358
DB 290 CPNHFV-VDQFFCVACPPDKMEVD-KHGLKKCEPCGGLCPACGCTGSG--SRQYVDS 345
QY 359 ANIOFACCKTIFGLALPESFDDPASNTPLOPELOVETLEETIGYLIYIAMPWS 418
DB 346 SNIDGFVACTKILGLDPLITGLANVPWPKI PALDPEKLVNRYRTREITGYLNIQSWPEH 405
QY 419 LPDLSVFNQLQVIRIRLIHNGAYS-LTLOGIGISWLGRLSELGSLALHNNHTLCV 477
DB 406 MENEVEFSNLTITIGRSLYNKGFSILMKLVNTSLGRLSKETISAGRYVIANQOLCYH 465
QY 478 HTVPMDQLFRNPQALLHTA-NRPBDCVGEGLACHQLCARGCHWGPGPTQVNCVSQFLR 536
DB 466 HSLNMTLLRLRPSERLIDIKYDRPLGECLAEKVCDDPLCSSGGCGPGGQCLCRNYSR 525
QY 537 GQCECEBRVLOGLPREVNAHCLPCHPEQCPQWGSYTCGPEADOCYACHYADPPPC 596
DB 526 EGVCTYHGNPLQGEEREFVHAQCFSCHPCLPMEGTSTCNGSGSDACARCAHFFDGPIC 585
QY 597 VARCPGKVPDLSTYMPIMKPFDEGACQPCPINCTHSC--VDLDDKCGPAEQRASPLTIS 654

```

Db      586  UNSCPHGLG--AKGPIYKPDPAQNECPCHENCTOGCNGPELQCLQAEVLSKPHLV 643
Qy      655  USAVGILLVVLGVFGILLIKRROOKIR-KYTMRLLOETELVPLTPSGAMPQAOQR 713
Db      644  IAVTVG--LAVIIMILGSGFLYMGRRIRIONKRAMRYLERGESIFPLDP--EKANKVLAR 700
Qy      714  ILKTELKRVKVLGSGAGFYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYV 773
Db      701  IFKETEIRKLKVLGSGVGVTHKGIMIEGESIKIPVCIKVIEDKSGQSFQAVTDHMLA 760
Qy      774  MAGGSPVSRLLGICLTSTVQLVQLMPLVCGCLLDHVENGRGSLGQOLLWMCQIACM 833
Db      761  VGSIDHAHIVRLGLCPSSSLQVLTVQYPLSLDLHVQHHTGLPOLLWVGQIACM 820
Qy      834  SYLEDVRLVHDLAARNVLVKS PNHVKTITDFGLARLDDIDETEVADGKVPKIMMALES 893
Db      821  YLEEHSHVYHDLALRNVLKSPGVQVADFGVADLLPPDDKQLHSAKPIKIMMALES 880
Qy      894  ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPEPCTIDVYM 953
Db      881  IHFKKYTHQSDVMSYGVTVWELMTFGAPYAGLRLAELPDLLEKGERLAPQICTIDVYM 940
Qy      954  IMVACMMIDSECRPRELVEFSRMDPQRPVNITONEDGPSPLDSTYRSLLEDD 1013
Db      941  VMVACMMIDENIRPTELANEFTRMADPPRYLTKAS--GGGTP--PAAEPSVLTITKE 997
Qy      1014  MGDVLDAEYLVLPQGFPCPDPAQAGGMVHRRHSSSTRSGCGDLTGLEPSEE----- 1068
Db      998  L-----GEABLEPEL-----DLDDLAEDEGLATS 1023
Qy      1069  -----EAPRPLAPSEG-----AGSDVFDGLGMAKGLSLPTH 1105
Db      1024  LGSALSLPTGTLTRPRGQSLSPSSGGMNOSLGRACLDASVLAGREGFSPRISLH- 1082
Qy      1106  PSPLORSSEPTVLPSETDGV---APL-----TC-----SPOE-----YVNOQDV 1145
Db      1083  PIPGR-----PASESSEGVGSELELOEKVSVCSRSSRSPRPRGSAVHSQHS 1135
Qy      1146  RPOPPSPREG-----LPAAPAGATLERAKTLSP-GKNGV-----KDYPAF 1187
Db      1136  LITVTLSPGLGEEDGNGVMPDTHLRGSSSREGTLLSVGLSVLTGEEDBD----- 1191
Qy      1188  GGAVENTEYLTPGGAAPQPHPP 1210
Db      1192  -----EEYEVNMRKRGRSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Insecta: Metazoa: Arthropoda: Mandibulata: Pancrustacea; Hexapoda;
OC Insecta: Pterygota: Neoptera: Endopterygota: Diptera: Brachycera;
OC Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 11
RX SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RL that several genetically defined classes of alleles cluster in
RP subdomains of the receptor protein.";
RP Genetics 137:531-550 (1994).
RP (2)
RP REVISIONS.

```

```

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livnen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RL hormone binding and kinase domains.";
RL Cell 40:599-607 (1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RL Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101 (1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Specific levels of EGF receptor signaling during photoreceptor
RL development in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144 (1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoisinger R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorelli J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Syvakas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasearan D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=8517938; PubMed=2983232;
RA Madaworth S.C., Vincent W.S. III, Bildeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
RP

```

RT growth factor receptor";
 RN Nature 314:178-180(1985).
 (8)
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/Flb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 (9)
 RN REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Pettimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the Drosophila
 RT EGF receptor.";
 RL Cell 89:13-16(1997).
 CC - FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MARK PATHWAY INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEKOSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC - ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS, IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AF052754; AAC08536.1; -;
 DR EMBL, AF052753; AAC08536.1; JOINED.
 DR EMBL, AF052754; AAC08535.1; -;
 DR EMBL, AF052752; AAC08535.1; JOINED.
 DR EMBL, K03054; AA551462.1; -;
 DR EMBL, K03417; AA551460.1; -;
 DR EMBL, K03416; AA50965.1; -;
 DR EMBL, K03418; AA551461.1; -;
 DR EMBL, AF109077; AAD26134.1; -;
 DR EMBL, AF109078; AAD26132.1; -;
 DR EMBL, AF109082; AAD26132.1; JOINED.
 DR EMBL, AF109078; AAD26133.1; -;
 DR EMBL, AF109078; AAD26133.1; JOINED.
 DR EMBL, AF109084; AAD26133.1; JOINED.
 DR EMBL, AF109079; AAD26130.1; -;
 DR EMBL, AF109081; AAD26130.1; JOINED.
 DR EMBL, AF109079; AAD26131.1; -;
 DR EMBL, AF109083; AAD26131.1; JOINED.
 DR EMBL, AF109080; AAD26135.1; -;
 DR EMBL, AE003454; AAF45732.1; -;
 DR EMBL, X02283; CA226157.1; -;
 DR EMBL, X78920; CA555523.1; -;
 DR EMBL, X78918; CA555521.1; -;

DR EMBL, X78919; CA555522.1; -;
 DR PIR, A00640; GQFE.
 DR HSPF, P1362; IFGK.
 DR FlyBase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR0109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
 KM Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KM Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426
 FT DOMAIN 31 868
 FT TRANSMEM 869 889
 FT DOMAIN 890 1426
 FT DOMAIN 938 1198
 FT NP_BIND 944 952
 FT BINDING 971 971
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 Query Match 27.9%; Score 1891; DB 1; Length 1426;
 Best Local Similarity 32.4%; Pred. No. 66-95;
 Matches 461; Conservative 185; Mismatches 440; Indels 338; Gaps 41;
 24 QVCTGDMKRLRPAPEHRLDMLRLHLYGCGVVGQNLLETYPVPT-NASIFLDIOEYVG 82
 100 KICIGTKRLSVSPSKHHRNLRDRTNCTVVDGNTKLTLPNENLDLSPFDNIREVVG 159
 83 YVLIHNVQVQVPLRLRLVRGTQLF-----EDNYALALDNGDPLANTTPTVGASPGCL 137
 160 YVLIHNVQVQVPLRLRLVRGTQLF-----EDNYALALDNGDPLANTTPTVGASPGCL 203
 138 RELQRLSLTEILKGVLIQRNPOLCYQDTILMKDIFPKNNQALTLIDNRSRACHPSG 197
 204 YVLEIPDLRLDVNLGVGFHNNYNLGMRTIQSEIVSNGTDAYVYNDFTAPRECPKCH 263
 198 WCKGRKCGSESEDDQSLRTVYACAGCA--RCKGPLROYIKANSKFIQITELKHSCLAC 255
 264 SCTHG-CWGEGRKNCQKSKLTCSPOCAGRCYGPAPRECCHLFCAGGCTGPTQKDCIAC 322
 256 LHFNSGICELCPALVTYNTDTPESMPREGRYTFGASCYACAPYNYLSTVGSCTLYC 315
 323 KNFFPEANYSKECPMRKRNKPTTYVLETNPEKVAVGACVACEP-CHLLRNGACVRCG 381
 316 PLHNGEVTAEADGTQRCCKSPCARVCYGLGMEHLREVAVTSANIQEFAGCKYFGSLA 375
 382 PDKKNDKGE-----CVPNCGPCPTKCPGVTLH-----AGNIDSRNCTVIDGNIR 428
 376 PLPESFDG--DPASNTA-----PLQPELOVPELTLEETGYLYISAWDSLPDLSVPON 427
 429 ILDQTFSGQDYAVAYNTMGPRYIPDPPEERVSVTSKEITGLNTEGTHPQFRNLSYFKN 488
 428 LQVIRGRILHNGAY-SLTLOGIGISWLGRLSRLSGSLALIHNTHLCFVHTVPDQLF 486
 489 LETIHGRQIMESMFALAIIVKSSLSLEBRNLKQSSGSVVQHNRDLCYGVNIRPAIO 548
 487 RNPQALLHTANRPDECEYGBGLACHQLCARHCWGPGFTQCVNCSQPLRGCECYEECV 546
 549 KEPEQKVVWVNNELRADLCEKNKTTICSDQCNEDGCGAGTQDLTCKNFNFNGCTIADCGY 608
 547 LQGLPREYVNAHNCPCRPBECQPNQSGVTCFPEADQCAAHYDPPRCVARNP----- 601

[illegible]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Milnen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Ciltenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M10066; AAA48763.1; ALT_INIT.
CC PIR, A00643; TVCHLV.
CC PIR, B00643; TVPVLV.
CC HSSP, P11362; 1E6K.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC PRINTS: PR001009; TYRKINASE.
CC ProDom: PD000001; Euk_Dkinase; 1.
CC SMART, SM00219; TyKc; 1.
CC PROSITE, PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE, PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE, PS50011; PROTEIN KINASE DOM; 1.
CC Transfaser: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC FT DOMAIN 132 399 PROTEIN KINASE.
CC FT NP BIND 138 146 ATP (BY SIMILARITY) .
CC FT BINDING 165 165 ATP (BY SIMILARITY) .
CC FT ACT SITE 257 257 BY SIMILARITY.
CC SQ SEQUENCE 634 AA: 70891 MW: 870533ABED01FCC CXC64;

Query Match 25.8%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1,1e-87;
Matches 370; Conservative 79; Pident 135; Indels 123; Gaps 17;

QY 587 CAHKKDDPPFCARPCSGVKPDLSTMPIMKPFDEGACQCPICINCHTSCVDLDDKCPAEQ 646
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 3 CAHPTDGHCVKCAPAGVAGENDTL-WKXADANAACVQLCHPNTTRCGKPGLEGP--- 58
QY 647 RASEPLTSSIVSAV-VGILLVVIAGVFGILIKRQOKIRKYMRRLLOETELVEPLTPSGA 705
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 59 NGSKTPSIASAAGVGGLLCVVYGIGLGLYLR-R-HYRKRLRLLOERELVEPLTPSGE 117
QY 706 MPNQAQMRILKETELRKVKVLGSAFGTVYKGIWIPGENVYKIPVAIKVLRNTSPYANK 765
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 118 APNQAHLRIKETEFKKVKVLGSAFGTVYKGIWIPGEEKYKIPVAIKELRENTSPYANK 177
QY 766 EILDEAYVMAGVSPYSLRLGILCTLTSTVQLVTOLMPYGLLDHVENRGRLSGODLLNW 825
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 178 EILDEAYVMASVDNPHVCRLLGILCTSTVQLITOLMPYGLLDVIRBHKNIQSQYLLNW 237
QY 826 CMOIAKGSYVEDVRLVHRLDAAANVLYKSPNHKRTDPGRLALDIDETEVADGGKVP 885
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 238 CVQIAKGNVYIEERLVRHRLDAAANVLYKTPQHKILTDPGAKLGLADEKRYHAEGGKVP 297
QY 886 IKWMALESLIRRFTHOSDWSYGVTVWELMTFCAPKPYDGI PAEIEPDLLEKGRLLPQP 945

```

```

Db      298 IKMALLESILRIRYTHQSDVWSYGTVMELMTFGSKPTDGIPTASISSVLEKGERLPQPP 357
Qy      946 ICTIDVYIMWKCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db      358 ICTIDVYIMWKCMWIDADSRPKFRELIAEFKMARDDPRYLVIQGDERMHLPSPTDSKF 417
Qy      1005 YRSILEDMDMDLVNAEYVLPQCGFPCPDPAFGAGVNHHRHSSSTRSGGDLTLGLE 1064
Db      418 YRTLMEEEDMDIVDADEYLVPHQGF-----NSPST----- 449
Qy      1065 PSEERAPSP-----APSEGASDVFDGLGMAKGLQSLPTDPSPLQRYSEDPTVP 1119
Db      450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHFPAVEDSFVQRYSSDPICN 495
Qy      1120 LPSET--DGYVAAPLTCSQPEPVYVNDVPRPQPSRECEPLPAARDAGATLERAKTUSPK 1177
Db      496 FLEESIDDGFL-----PAPEYVNO--LMPKKPS-----TAMVQ 526
Qy      1178 NGVYKDVAF-----AFGAVENPEYLTPOGGAAPQHPHPPAPSPAFDNLVY 1222
Db      527 NQIYNNISLTAISKLPMDSRVONSHSTAVDNEPYL-----NTNQSBLAKTVFESSPY 578
Qy      1223 WDO-----DPEP-----RGAPSTFKGTPTAENPEYLTGLDVP 1254
Db      579 WIQSGNHQIINLDNPYQODFLPNETKPNGLLKVPAENPEYLRVAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
ID ERBB_AVIER
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=9685;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78 (1983).
RN (2)
RN SEQUENCE OF 1-152 FROM N.A.
RP MEDLINE=84223957; PubMed=6328658;
RA Debutre B., Henry C., Benajissa M., Biserte G., Claverie J.-M.,
RA Sauls S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459 (1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as there is no content in it
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).

```

```

DR EMBL: K02006; AAA42394.1; ALT_INT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TVTQH.
DR HSSP: P11362; IFGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD067450609 CRC64;

```

```

Query Match      25.1%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 3,3e-85;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

```

Qy      587 CAHYKDPFCVYARCPGVPDLSTYPIWKFPDEGACQPCPINTHSCVDLDDKCCPAQ 646
Db      3 CAHFIDGPHCVYACGAVLGENDTL-VKRYDANAVCOLCHPNCRTGCGPQLEGGP--- 58
Qy      647 RASPLTISVASAV-GILVVVLGVVFGILIRKROOKIKRTYTRRLQELVELPLTPSQA 705
Db      59 NGSKTPSIAAGVGGLLLVVGLIGLYLRRL-HIVKRTLRRLQELVELPLTPSSE 117
Qy      706 MPNOKMRLKETELRKTVKVGSAFGTVYKIMIPDGENVKIPAIKIVKRENTSPKAK 765
Db      118 APNOKHLILKETEKTKVKVLSGAFGTIYKGMVPEBEKVIPIAIKELRATSPKAK 177
Qy      766 EILDEAYVMAGVSPYVSRLLIGICTSTVQVLVTQMLPYGCLLDHVRNKRGLSGDLLNW 825
Db      178 EILDEAYVMAGVSNHVCRLIGICTSTVQVLVTQMLPYGCLLDYIREHKNIGSOYLNM 237
Qy      826 CMQIAKGSYLEDVLVNRDLAARVVLKSPNHVITPQGLARLIDIDETEVHADGKVP 885
Db      238 CVQIAKGNVLEERLVHRDLAARVVLKTPQHVKITPQGLAKLGADEKEYHAEGKVP 297
Qy      886 IKMALLESILRRRPTHQSDVWSYGTVMELMTFGAKPYDGIPTAREIPDLLEKGERLPQPP 945
Db      298 IKMALLESILRIRYTHQSDVWSYGTVMELMTFGSKPTDGIPTASISSVLEKGERLPQPP 357
Qy      946 ICTIDVYIMWKCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db      358 ICTIDVYIMWKCMWIDADSRPKFRELIAEFKMARDDPRYLVIQGDERMHLPSPTDSKF 417
Qy      1005 YRSILEDMDMDLVNAEYVLPQCGFPCPDPAFGAGVNHHRHSSSTRSGGDLTLGLE 1064
Db      418 YRTLMEEEDMDIVDADEYLVPHQGF-----NSPST----- 449
Qy      1065 PSEERAPSP-----APSEGASDVFDGLGMAKGLQSLPTDPSPLQRYSEDPTVP 1119
Db      450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHFPAVEDSFVQRYSSDPICN 495
Qy      1120 LPSET--DGYVAAPLTCSQPEPVYVNDVPRPQPSRECEPLPAARDAGATLERAKTUSPK 1177
Db      496 FLEESIDDGFL-----PAPEYVNO--LMPKKPS-----TAMVQ 524
Qy      1178 NGVYKDVAF-----AFGAVENPEYLTPOGGAAPQHPHPPAPSPAFDNLVY 1222
Db      527 NQIYNNISLTAISKLPMDSRVONSHSTAVDNEPYL-----NTNQSBLAKTVFESSPY 574
Qy      1219 NLVYWDODPPRPGAPRPTFKGTPTAENPEY 1248

```


Db 575 SSPYWIOSGNHQ-----INLDNPDY 594

RESULT 14
ERBB_AVIEU STANDARD: PRT: 540 AA.

AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 11, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN v-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Traitor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells."
RT Mol. Cell. Biol. 6:1751-1759(1986).
RL 1;
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL_M13179; AAA42401.1; -
DR PIR; A25321; TYFVEB.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase_1.
DR ProDom; PD000001; Euk_pkinase_1.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR TRANSFERASE; PS50011; PROTEIN_KINASE_DOM; 1.
KM Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KM Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 24.0%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 2.6e-81;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 587 CAHYKPPPCVAVACPSGVKDLSTYMPKPPDEBGAQCPPIINCTHSVDLDDKCPAQ 646
DB 3 CAHFIDGPHCVKACPGAVLQENDTL-VMKYADANAVQQLCHPCTRGCKPGLEGCP--- 58
QY 647 RASPLTSISAVV-GILLVVVLGVFGILLKRRQOKIRKXTMRRLLOETLVEPLTPSGA 705
DB 59 NSKSTPSIAAGVGGILLVVLGVFGILLKRRQOKIRKXTMRRLLOETLVEPLTPSGE 117
QY 706 MPNOAQRILKETELRKVYLSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 765
DB 118 AENQALRIILKETEFKVKVYLSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 177
QY 766 ELIDVAVYVAGVSPVSRLLGLCTSTVQVYQVLPYGLLDHVENRGLSGDILLN 825

Db 178 ELIDVAVYVAGVSPVSRLLGLCTSTVQVYQVLPYGLLDHVENRGLSGDILLN 237

QY 826 CMQIAKGSYLEDVRLVRLDAARVLYKSPNHYITDQGLARLIDIDETEHNAQGVK 885
DB 238 CVQIAKGNVLYEENHVARDLAARVLYKSPNHYITDQGLARLIDIDETEHNAQGVK 297
QY 886 IKMNALESILRRPETHOSDVMSYGVYTWELMTFGAKPYDGIIPAREIPDLEKGERLP 945
DB 298 IKMNALESILRRPETHOSDVMSYGVYTWELMTFGAKPYDGIIPAREIPDLEKGERLP 357
QY 946 ICTIDVYVYKCMWIDSECPRRRELVSERNAQDFQFVQVQ-NEDLGASPLDSTF 1004
DB 358 ICTIDVYVYKCMWIDSECPRRRELVSERNAQDFQFVQVQ-NEDLGASPLDSTF 417
QY 1005 YRSLIEDDMGDLVDAEYLYVPOQGFPPDPDAPAGGVNHRHSSSRSGGDLTLGLE 1064
DB 418 YRSLIEDDMGDLVDAEYLYVPOQGFPPDPDAPAGGVNHRHSSSRSGGDLTLGLE 449
QY 1065 PSEEGAPRSPV-----APSEGAGSDVFDGDMGAKGLQSLPTHPDPSLPDRYSEDPTVP 1119
DB 450 -----SRTPLLSLATSNSNATNCIDRNG-----H----- 476
QY 1120 LPSTDGIVAPLTCSPQPEYVNPDPVNPQPSPEGGLPAARPAQAT-LERAKTLLSPGN 1178
DB 477 -PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQVYVYSLTATSKLPIDSRYN 527
QY 1179 GVAVDVAFAGVAVENPEYL 1197
DB 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR_CHICK STANDARD: PRT: 703 AA.

ID EGFR_CHICK
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RT Ulrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mouse cells, and differential binding of EGF and transforming growth factor alpha."
RT Mol. Cell. Biol. 8:1970-1978(1988).
RL 1;
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF, AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to dimerization, internalization of the EGF-receptor complex, induction of the tyrosine kinase activity, stimulation of cell DNA synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).


```
CC or send an email to license@eb-sib.ch).
DR EMBL: M20386: AAA48760.1: -
DR InterPro: IPR000494: EGFR_L_domain.
DR InterPro: IPR000719: Euk_kinase.
DR InterPro: IPR002174: Furin-like.
DR InterPro: IPR001245: Tyr_kinase.
DR Pfam: PF00757: Furin-like; 1.
DR Pfam: PF01030: Recep_L_domain; 2.
DR SMART: SM00261; FU; 4.
DR PROSITE: PS00107: PROTEIN_KINASE_ATP_PARTIAL.
DR PROSITE: PS00109: PROTEIN_KINASE_TYR_PARTIAL.
DR PROSITE: PS50011: PROTEIN_KINASE_DOM_PARTIAL.
DR Transmembrane: Glycoprotein: Receptor: Signal: Transferase;
KW Tyrosine-protein kinase: ATP-binding; phosphorylation.
FT SIGNAL 1 30
FT CHAIN 1 30
FT DOMAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT TRANSMEM 31 654 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA: 77427 MW: 492DE11B735A690 CRC64;

Query Match 22.6%; Score 1532; DB 1; Length 703;
Best Local Similarity 43.3%; Pred. No. 7e-76;
Matches 306; Conservative 112; Mismatches 261; Indels 28; Gaps 12;

QY 8 RMGLLALLPPGAA-----STOVCTGTDMLKRLPASPETHLDMLRHLYOGCQVQGNLE 61
DB 13 RGAAVLVLLVLLGVALCSAVEEKVCQGTNNKLTQLGHVEDHPTSLQRMYNKCEVLLSNLE 72
QY 62 LTYLPTNASLAFLODIOEVGVVLIANOVROVPLQRIYRGVQLFEDNYALAVLDNGD 121
DB 73 ITTYEHRDLTFELKTIQEVAVGVLIALNMVDVILENIQIRGNVLYDNSFALAVLSNYH 132
QY 122 PLNNTTPVTGASPGGLARELQRLSLTEILKGVLLIQRPOLCYQDTIILWKDI FHKNNQAL 181
DB 133 -MNKTO-----GLRELPMKRLSEIILNGVKSINNPKLCNMDTVLWMDIIDSRSK-PL 182
QY 182 TLID-TIRSRACHQSPKWCCKSRCKGSESDCQSLTRTVCAAGCA-RCKGPLPOYIKANS 239
```

```
DB 183 TVLDPASNLSSCPKCHPNCTEDHCGAGEQNCQTLTKVICAQOCSGRCRGKVPSPDCCHNQ 242
QY 240 KFIQITELKXSDCLACLFHNHSGICELHCPALVYNTNDTFESMNPREGRYTFGASCYTAC 299
DB 243 CAACCTGPRESDCLACRFRDADCTKDCPLVLYNPTTYQMDVNPBEGKYSFGATCVNRC 302
QY 300 PNYLSTDVGSCTLVCPPLHNOEYTAEDQTORCEKSKPCARVCYGLGMEHLREYAVTSA 359
DB 303 PHNYVYTDHGSVCVSCNTDTEYV-BENCVRKCKKCDGLCSKVCNGICIGELKGLISNAT 361
QY 360 NIOEFAGCKI FGSIAFLPESFDGDPASNTAPLOPEOLQVEFTELEITGYLYISAMPDSL 419
DB 362 NIDFFKCTKINGDVSLPVAFLDPAFTKTLPLDPKLDVFRVWEISGFLLIQMPDNA 421
QY 420 PDLSPQNLQVIRIRILHNGAYSLTQLOGISWLGRLSRLSGSLAIHHNTHLCFVHT 479
DB 422 TDLVAFENLEIRGRTRQHGQYSLAVNVLKIQSLGRSLKEISDIDAIMONKMLCVADT 481
QY 480 VPMQDLFRNPQALLHTANRPEDECVSGELACHQLCARGHCGMGRPTQCVCNCSOPLRGQE 539
DB 482 MNWSSLFATQSQTKIIONRKNDCTADRHVCDPCLCSVCGMCGRPHCFSCRFPSRQKE 541
QY 540 CVBECRVLOGLPREYVNAHCLPCHPECGPONG---SVTCFGEADQCVACAHYKDPFC 596
DB 542 CVKOCNIIQGBRPFERDSKCLPCHSECLVONSTAYNTTCSGPGPDHCKCAHFDGPHC 601
QY 597 VARPSPGVKPRDLSYMPIMKFPDEBGAQCPPIINCHTSCVDLDDGDCRPAEQASPLTYSV 656
DB 602 VRACPAGVLGENDTL-VKRYADANAVCOLCHPNCRCGCKGGLCCP---NCKSTPSIAA 657
QY 657 AVV-GILLVVVLGVVFGILIRKROOKIRKYMRRLQETELVEPLTP 702
DB 658 GVVGGLCLTVVVGIGIGLYLRR-HIVRKRLRLRLORELVEPLTP 703
```

Search completed: July 22, 2003, 08:45:06
Job time : 20.2304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(Without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-210-224-12

Perfect score: 6789
Sequence: 1 MELALACRGLALLPFGA.....TRKGTPTANPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioid:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6187	91.1	1259	6	018735
2	3079	45.4	1209	11	09QX70
3	3050	44.9	1210	11	09BP98
4	2659	39.2	1165	13	09YH40
5	2635.5	38.8	1137	13	09W6F6
6	2241	33.0	1128	13	P79754
7	1961.5	26.0	1433	5	09B1H9
8	1766.5	26.0	419	4	09UK79
9	1739	25.6	367	11	08R2X1
10	1720	25.3	729	15	086712
11	1718	25.3	567	15	086714
12	1697.5	25.0	412	4	08WYV0
13	1653.5	24.4	962	15	064895
14	1645	24.2	545	15	085468
15	1441.5	21.2	655	11	09WVF5
16	1425.5	21.0	643	11	09BRV6

17	1217	17.9	1193	5	09Y1X8	09Y1X8 ephedratia f
18	1161.5	17.1	1368	5	023821	023821 caenorhabdi
19	1155	17.0	1717	5	026566	026566 schistosoma
20	1063	15.7	527	13	090836	090836 gallus gall
21	971.5	14.3	599	13	09PSH2	09PSH2 gallus gall
22	936.5	13.8	478	11	09RSE0	09RSE0 rattus norv
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.1	176	11	0923V5	0923V5 rattus norv
25	806.5	11.9	346	13	P11776	P11776 xiphophorus
26	778	11.5	435	5	08SZM1	08SZM1 drosophila
27	754.5	11.1	311	13	099162	099162 xiphophorus
28	712.5	10.5	1362	13	09PVZ4	09PVZ4 xenopus lae
29	683	10.1	1671	5	09NIV5	09NIV5 biophidari
30	658.5	9.7	1368	13	08UW85	08UW85 paralicthy
31	648	9.5	331	4	09BUD7	09BUD7 homo sapien
32	643	9.5	1418	13	093457	093457 scophthalmu
33	626.5	9.2	1472	5	09U5A8	09U5A8 bombyx mori
34	626	9.2	1358	13	073798	073798 xenopus lae
35	622.5	9.2	1369	13	08UW86	08UW86 paralicthy
36	615	9.1	149	6	09BG66	09BG66 oryctolagus
37	611	9.0	1412	13	08UW84	08UW84 paralicthy
38	602	8.9	987	11	091YMO	091YMO mus musculu
39	599	8.8	987	11	099MR2	099MR2 mus musculu
40	595	8.8	1371	11	09QVW4	09QVW4 rattus norv
41	592.5	8.7	2144	5	09VD94	09VD94 drosophila
42	590.5	8.7	1418	13	08UW83	08UW83 paralicthy
43	587.5	8.7	1036	4	007912	007912 homo sapien
44	587.5	8.7	1055	11	054967	054967 mus musculu
45	586.5	8.6	1091	4	09UWQ4	09UWQ4 homo sapien

ALIGNMENTS

RESULT 1
018735
ID 018735 PRELIMINARY; PRT; 1259 AA.
AC 018735;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Tokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland."
RL Submitted (0CT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; 1FCG.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 91.1%; Score 6187; DB 6; Length 1259;
Best Local Similarity 91.0%; Pred. No. 0;
Matches 1147; Conservative 39; Mismatches 68; Indels 6; Gaps 2;

QY 1 MELALALRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLLDMLRLYLGGCCVVGQNL 60
DB 1 MELALALRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLLDMLRLYLGGCCVVGQNL 60
QY 61 ELTVLPNTASLSFLQDIQEVGVYVLAHNOVQVPLQRLIVRGTLFEDVYALAVLDNG 120
DB 61 ELTVLPNTASLSFLQDIQEVGVYVLAHNOVQVPLQRLIVRGTLFEDVYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNOLA 180
QY 121 DPLEGGIPAPGAAGGJRELQRLSLTEILKGGVLIQRNPOLCHQDTILMKDVFHKNQOLA 180
DB 121 DPLEGGIPAPGAAGGJRELQRLSLTEILKGGVLIQRNPOLCHQDTILMKDVFHKNQOLA 180
QY 181 LTLIDTRSRACHCSPEWCKSGRCMGESSEDCQSLTRVYCGGACRCGRLPOTIKANSK 240
DB 181 LTLIDTRSRACHCSPEWCKSGRCMGESSEDCQSLTRVYCGGACRCGRLPOTIKANSK 240
QY 241 FIGITELKHSDDLACLFHNSGICELHCPALVTYNTDFEESMPNPEGRTGASCVTACP 300
DB 241 FIGITELKHSDDLACLFHNSGICELHCPALVTYNTDFEESMPNPEGRTGASCVTACP 300
QY 241 AAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDFEESMPNPEGRTGASCVTACP 300
DB 241 AAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDFEESMPNPEGRTGASCVTACP 300
QY 301 YNVLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVYSAN 360
DB 301 YNVLSTVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVYSAN 360
QY 361 IOEPAGCKKIFGSLAFPESEFDPDASTAPLQEPOLQVFEFTLEITGYLYISAMPDLP 420
DB 361 IOEPAGCKKIFGSLAFPESEFDPDASTAPLQEPOLQVFEFTLEITGYLYISAMPDLP 420
QY 421 DLASFQNLQVIRGRILHNGAYSLTLQGLGIMLGRLSRELSGSLALHNTHLCEFYTV 480
DB 421 DLASFQNLQVIRGRILHNGAYSLTLQGLGIMLGRLSRELSGSLALHNTHLCEFYTV 480
QY 481 PMDOLFNNPHQALHTANPREDECVGBGLACHOLCARGHCGPFGTCVNSQFLRGQEC 540
DB 481 PMDOLFNNPHQALHTANPREDECVGBGLACHOLCARGHCGPFGTCVNSQFLRGQEC 540
QY 481 PMDOLFNNPHQALHTANPREDECVGBGLACHOLCARGHCGPFGTCVNSQFLRGQEC 540
DB 481 PMDOLFNNPHQALHTANPREDECVGBGLACHOLCARGHCGPFGTCVNSQFLRGQEC 540
QY 541 VEEGRVIOGLPREVYNARHCLPCHPECOPONGSVTTCREPRADOCVACHYDDPPFCVAC 600
DB 541 VEEGRVIOGLPREVYNARHCLPCHPECOPONGSVTTCREPRADOCVACHYDDPPFCVAC 600
QY 540 VEEGRVIOGLPREVYNARHCLPCHPECOPONGSVTTCREPRADOCVACHYDDPPFCVAC 599
DB 540 VEEGRVIOGLPREVYNARHCLPCHPECOPONGSVTTCREPRADOCVACHYDDPPFCVAC 599
QY 601 PSQVXPDLISYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAPORASPLTISVSAVVG 660
DB 601 PSQVXPDLISYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAPORASPLTISVSAVVG 660
QY 600 PSQVXPDLISYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAPORASPLTISVSAVVG 659
DB 600 PSQVXPDLISYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAPORASPLTISVSAVVG 659
QY 661 ILLVVLGVVFGILLIKRQCKIRKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 720
DB 661 ILLVVLGVVFGILLIKRQCKIRKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 720
QY 660 ILLVVLGVVFGILLIKRQCKIRKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 719
DB 660 ILLVVLGVVFGILLIKRQCKIRKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 719
QY 721 RKVAVLGSAGFAGTYKGIWIPDGENVKIPVAIKYLRENTSFKAKELIDELVYVWAGVSP 780
DB 721 RKVAVLGSAGFAGTYKGIWIPDGENVKIPVAIKYLRENTSFKAKELIDELVYVWAGVSP 780
QY 720 RKVAVLGSAGFAGTYKGIWIPDGENVKIPVAIKYLRENTSFKAKELIDELVYVWAGVSP 779
DB 720 RKVAVLGSAGFAGTYKGIWIPDGENVKIPVAIKYLRENTSFKAKELIDELVYVWAGVSP 779
QY 781 YVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRGLSGODLLNMCMQIAKMSYLEDDR 840
DB 781 YVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRGLSGODLLNMCMQIAKMSYLEDDR 840
QY 780 YVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRGLSGODLLNMCMQIAKMSYLEDDR 839
DB 780 YVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRGLSGODLLNMCMQIAKMSYLEDDR 839
QY 841 LVHRDLAARVVLKSPVHVKITDFGLARLDIDETEHADGKVPIMKMALESILRRFT 900
DB 841 LVHRDLAARVVLKSPVHVKITDFGLARLDIDETEHADGKVPIMKMALESILRRFT 900
QY 840 LVHRDLAARVVLKSPVHVKITDFGLARLDIDETEHADGKVPIMKMALESILRRFT 899
DB 840 LVHRDLAARVVLKSPVHVKITDFGLARLDIDETEHADGKVPIMKMALESILRRFT 899
QY 901 HOSDVMASGYTVVMTLFGAKPYDGIIPAREIPDLLEGEERLPORPCTIDVYIMVACMM 960
DB 901 HOSDVMASGYTVVMTLFGAKPYDGIIPAREIPDLLEGEERLPORPCTIDVYIMVACMM 960
QY 900 HOSDVMASGYTVVMTLFGAKPYDGIIPAREIPDLLEGEERLPORPCTIDVYIMVACMM 959
DB 900 HOSDVMASGYTVVMTLFGAKPYDGIIPAREIPDLLEGEERLPORPCTIDVYIMVACMM 959
QY 961 IDSECRPRFRELVESEFSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDDMGDLVDA 1020
DB 961 IDSECRPRFRELVESEFSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDDMGDLVDA 1020
QY 960 IDSECRPRFRELVESEFSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDDMGDLVDA 1019
DB 960 IDSECRPRFRELVESEFSRMARDPQRFVYIQNEDLGASPLDSTFYRSLLEDDMGDLVDA 1019
QY 1021 EFLYLVPOGGFCPPDAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
DB 1021 EFLYLVPOGGFCPPDAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080

DB 1020 EFLYLVPOGGFCPPDAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1079
QY 1081 AGSDVFPDGLGMAKAGLQSLPTHDDPELQRYSDPTVPLPSEITDGYAPILTCSQPERYV 1140
DB 1080 AGSDVFPDGLGMAKAGLQSLPTHDDPELQRYSDPTVPLPSEITDGYAPILTCSQPERYV 1139
QY 1141 NQPDVPRQPPSPRSGPLPAARPAATLER-----AKTILSPKNGVYKDVAFGAVENPE 1195
DB 1140 NQPDVPRQPPSPRSGPLPAARPAATLER-----AKTILSPKNGVYKDVAFGAVENPE 1199
QY 1196 YLTPQGAAPQPHPPAFSPAFLDNLVYWDODPPERGAPPSFKGTPTAENBEYLGLDVPV 1255
DB 1200 YLTPQGAAPQPHPPAFSPAFLDNLVYWDODPPERGAPPSFKGTPTAENBEYLGLDVPV 1259

RESULT 2
Q90X70 PRELIMINARY; PRT, 1209 AA.
AC Q90X70;
AD Q90X70;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258688; PubMed=2342466.
RA Pech L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RL encoded by an alternatively spliced transcript in normal rat tissue.";
RN Mol. Cell. Biol. 10:2973-2982(1990).
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Pech L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: M37394; AAF14008.1; -
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fu; 3.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FE87F6C18773 CRC64;

Query Match 45.4%; Score 3079; DB 11; Length 1209;
Best Local Similarity 49.5%; Pred. No. 1,4e-224;
Matches 632; Conservative 166; Mismatches 365; Indels 114; Gaps 26;

QY 3 LAALCRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLLDMLRLYLGGCCVVGQNL 61
DB 3 LAALCRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLLDMLRLYLGGCCVVGQNL 61

```

Db      45 LAAALCAAG-----GALEEKVCOGTSNRLLTQLGFEDHFLSLQRMFNCEVLGNLE 66
Qy      62 LYYLPTNASLSFLQDIQEVGVYLIANOVQVPLQRLIRVSTQLFEDNYALAVLDNGD 121
      67 IYVQNNYDLSFLKTIQEVAGVYLIAMNTERI:PLENLQIIRGNALYENTYALAVLSN-- 124
Qy      122 PLANTTPVVGASGGRLELOLRSLTEILKGVLIQNRPOLCYODTTIMKQIFPKNNOLAL 181
      125 -----YGVNTKGTRELPMRNLQELIGAVRFSNNPILCNMETTIQWRIV- QDVFLSN 175
Qy      182 TLIDTNRS-RACHPCSPMCKSGSRMGESSEDCOSLTRTVGAGCA-RCKGPLQYIKANS 239
      176 MEMDVQRHLTGCKCPSCPCNSCGRGEECNCKTKIICAQCCSRRCRBSRSDCCCHQ 235
Db      240 KFIIGTELXSDCLACLFHNHSGICELACPALVTYNTDFESMPNPEGRYTFGASCVTAC 299
      236 CAAGCTGPRESDCLVCHRFDEATCKDTCPRLMLYNPTTYQMVNPEGKYSFCATCYKCK 295
Qy      300 PYNVLTSDVSGCTVPLRNLNOEYTAEDGTQRCCKSPCARVYCYGLGMELREVRAYTSA 359
      296 PKNVYVTDHSGCRACGPDYEV-EEDGVSKCKCGPCPKVCNGIGIGFKDTLSINAT 354
Qy      360 NIQEFAGCKKIFGSLAFLPESPFGDPAASNTAPLOPOLQVFETLEETIGVLYISAMPDSL 419
      355 NIKHFKYCTAISGDHLILPVAFKGDSFTTRPRLDREBELIKTVKXELTGFLLIQAMPENW 414
Qy      420 PDLVSVFQNLQVIRGRILHNGAVSLTQGLGISWLGRLSRLSELQGLALIHNTLFCVHT 479
      415 TDLHAFENLEIIRGRTKHQGFSLAVVGLNITSLGRSLKEISDGOVITISGNNLTCYANT 474
Db      480 VWMDOLEFRPHQALHTANRPRECECGBGLAQOLCARHGCKWGPPTQVCNQCQFLRGCE 539
      475 IMWKLEFGTPNQKTKIMNRAEKDCATNHNVCNPLSSBECWGPBETDCCVSCCNVSRGHE 534
Qy      540 CVEECRVLQGLPREYVNAHRLCPCHPECOPONGSVTCFGEADQCAVCAHYKDPRECVAR 599
      535 CYDKCNILEGEPRERFENSECIOCHPECPLQTMNITCTGGRPNCKICAKAYVNGPCVKT 594
Qy      600 CPBGVCKPDLISYMPWKFPDEBEGACQPCPINCTHSCVDLDKCGPABORASP-LTSIVASV 658
      595 CPBGINGENNTL-VWKFADANNVCHLCHANCTYGACGPGIKGC--QQPEKPIPSIATGI 651
Qy      659 VGLILVNVUGVVFGL-LIKRQOKIRKYMRLLOTEVELPTBEGAMPNOQMKILKE 717
      652 VGLLFETIV-VALGIGLFMRRLQVKKRTLRLLOREVEPLTPSGEABNOAHLRLIXE 710
Qy      718 TELRKVKVLSGAGFTVYKGIWIPDGENYKIPVAIVLRENTSPKANKKELLDEAYVMAGY 777
      711 TEFKKIKVLSGAGFTVYKGIWIPDGENYKIPVAIVLRENTSPKANKKELLDEAYVMASV 770
Db      778 GSPVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNMCQIAKGMSTYLE 837
      771 DNPHVCRLILGICLTSTVQLITQMLPYGCLLDVREHKDNIQGYLLNMCQIAKGMSTYLE 830
Qy      838 DVLRIVHRDLAARNVLYKSPNHYKIDPGLARLLDIDETVHADGCVPIKMALESILRR 897
      831 DRLVLRHDLAARNVLYKTPQHVKITDFGAKLIGAEKEHYHAEGGCVPIKMALESILRR 890
Qy      898 RFTHOSDVMSYGVYVWELMTFGAKPYDGI:PAEIPDLLEKGERLPOPPICITIDVYIMWK 957
      891 IYTHOSDVMSYGVYVWELMTFGSKPYDGI:PAEISILKGERLPOPPICITIDVYIMWK 950
Qy      958 CMMIDSECRPRPRELVSEFSRMAPDQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDMD 1016
      951 CMMIDADRPRPRELILEFSKMAPDQRFVVIQ:QDERMHLPSPTDSNFRALMEEDMDMD 1010
Qy      1017 LVDAEYVLPQCGFCPPDAPAGAGMWHHRHSSSTRSGGGLTLGLEPSEEDAPRSPILA 1076
      1011 VVDADYEYLPQCGFC-----NSPST-----SRTPLL 1036
Db      1077 PSEGASGVDFDGLDGMGAAGLQSLPTHDSPLOFORSSEDPVLPSET--DGVVAFLTCS 1134
      1037 SLSANSN-----SSTVACINRNGSCRVADAFLOKYSSTPTSVLTEDNIDTFL----- 1086

```

```

Qy      1135 POEYVNOQDVPDRQPPSPREGPLPAARPAQATLERAKTSLSPKNGVVKVDFAFGAVENP 1194
      1087 PVPEYINQ-SVPRKPPAGSVGNPVYHNPRLP-----ARGDHLHQN--PHSNAYSNP 1135
Db      1195 EYL-TPQGAAPQHPHPAPSPAFDNLVYWDQ-----DP-----PERGAPSTF 1237
      1136 EYLNTAQ-----PTCLSSGFSSALWIQGSHQMSLDNDPDYQDFFPKKAPMGIF 1186
Qy      1238 KGTPTANPEYLGLDVP 1254
      1187 KG-PTANAEYLRVAP 1202
Db
Qy
Db
RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bpidermal growth factor receptor isoform 1.
GN BGRF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearseall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP [12]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel C., Pearseall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP [1]
RP Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RA EMBL; AF275366; AAG28045.1; -.
RA EMBL; AF275364; AAG28045.1; JOINED.
RA EMBL; AF275365; AAG28045.1; JOINED.
RA EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MG1:95294; Egfr.
DR InterPro; IPR000345; CyC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Btk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Btk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```

KM ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134840 MM; 62CD021C9DE32E18 CRC64;
Query Match 44.9%; Score 3050; DB 11; Length 1210;
Best Local Similarity 49.0%; Pred. No. 2.2e-222;
Matches 623; Conservative 168; Mismatches 371; Indels 110; Gaps 23;
11 LILALLPAGAA--STOVCTGTDMKLR.LPASPETHLMDLRHLQYGCQVAVQGNLETTYPTN 68
14 LILALLCAGAGALEEKVKYCGQTSNRLTQGTGFEDHFLSLQRMKNCEVAVLGNLETTYQRN 73
69 AASLSPFLODIOEVQGYLIANNOVRQVLRIRVGTQGFEDNYALAVLNGDPLANTTP 128
74 YDLSFLKTIQEVAGYVLIANTVERIPLENQIIRGNALYENTYALALISN----- 124
129 VTGASRQGLRELRSLTELKGVLIQRNPOLCYOTILMKD-----FKKNQDLATLI 184
125 -YGTNRGLRELPRNRIQETILIGAVRSPNPILCNMDTIQMRDLYQNVFMSNMD- 180
185 DTRSRACHPSPKSPKSCSGESSEDCQSLTRTVACAGCA-RCKGPLPQYIKANSKFEIG 243
181 -QSHPSCKPCDCSDSCPGSCMGGEENCOKLTIKICQCSHRGRGSPSDCCNOCAG 239
244 ITLKLKSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNREGRYTFGASCYTACPNY 303
240 CTGPRESDDLVCQKFODEATCKDTCPPLMLYNPTTYQMDVNPBEGKYSFGATCVKCKPRNY 299
304 LSTDVGSCTLVCPPLHNOEVTAEQSTQCEKSKCAVCCGLGMEHLREVRANTSANIQE 363
300 VVVDHSGCVACGPDYEV--EEDGIRKCKCKDGCRCRYCNGIGEGFKDTLSIATNIKH 358
364 FAGCKKIFGSLAF.PESFDDDPASNTAPLOPELOVETLEIYGYIYSAMPSLDLS 423
359 FKYCTAISGDLHLPVAFKQDSFTRTPLDRELEIKYKEITGFLLIQAMPNMTDLH 418
424 VFONLOVIRGRILHNGAYSLTQGLGSMGLSRLBELSGSLAIHNTHLCFVHTVPM 483
419 AFENLEIRGRITKOHQGFSLAVVGLNITSLGLRSLKEISDDVDIISGNRLCYANTINMK 478
484 QLFNRHQALHTANRPEDCEVGEGLACHOLCARGHGWGRCPTQVNCOSPLRQEEVE 543
479 KLTGTPQOKTKIMNRRKCKKCAVNHVNCNPLCSSEGGWGPBRCCVSCQNVSRRECEK 538
544 CRVLOGLPREYVNAARHCLPCHECOPONGSVTFCGPADOCVACAHYKDPFCVACRPSG 603
539 CNLIEGPRREFVENSECICQHPBELPQAMNITCTGRBPDNICOAHNITDGHCVKCTPAG 598
604 VKPDLSTWPIWKEPDEEGACQPCPINCTHSCVDLDDKGCRAEORASPLTISVAVGIL 663
599 IMGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQGEVWPSGPKIPSIATGIVGGL 657
664 VVVLGVVFGI-LIKRROOKIRKYMTRRLLOETELVEPLTSGAMPNOAKRIKETELRK 722
658 FIVV-VALGIGLEFRRRHIVRKTLRLDERELVEPLTSGEAPNOAHRIKLETFPKK 716
723 VKVLSGSAFTVYKGIWIPDGENYKIPVAIKVLRNTSPKANKKILDEAVYMAVGSPV 782
717 IKVLGSAFTVYKGLMIPEGEKKIIPVAIKELREASPRANKKILDEAVYMAVSDPHV 776
783 SRLIGCLISTVOLVTOIMPYGLLDHVRNRLSGQDLLNMCQIAKGSYLEDLVLY 842
777 CRLLGICLSTVOLITQMLPYGCLLDVREHKDNGISQYLLNMCVQIAKGNLYLEDRLV 836
843 HRDLAARNVVKSPNHKIDTDFGLARLDIDETEVHADGGVPRKMMALSSILRRRTHQ 902
837 HRDLAARNVVKTPOHAKITDFGLAKLGAEEKYHAEAGGVPIKMMALSSILHRIYTHQ 896
903 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPRICTIDVYIMWKCMID 962
897 SDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGRLLQPPRICTIDVYIMWKCMID 956
963 SEGRPRRELVSSESRAPDPORVVQ-NEDLQSPAPLSTFYRLLEDDEDDMDLVAE 1021

Db 957 ADSRPKRELELFESKMARDPORYLIQGDREMLPSPDTSNFYRALMDEEDMEDVVDAD 1016
Qy 1022 EYLVPOGFEFCPPBPAPAGCMVHHRSSSTRSCGGDLTGLRESEEAAPSLAPSEGA 1081
Db 1017 EYLVPOGFE-----NSPST-----SRTPLLSSLSA 1042
Qy 1082 GSVFDDDLGMAKGLQSLPTHDPSPLQYSEDPVPLPSET--DGVVAPLTCSPQPEY 1139
Db 1043 TSN---NSTIVACINRSGSRVXKEDAFLOQYSSDPTCAVVEDNIDDAFL-----PVPEY 1092
Qy 1140 VNOQVAPQPPSPREGPLPAPAPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYL-T 1198
Db 1093 VNO-SVPRKPAQSVQNVVYHNP.LHP-----APGRDLHYQN--PHSNVGNPEYLYNT 1141
Qy 1199 POGGAQPHPPAPFAPFAPNLYWMD-----DP-----PERGAPSTFKGTP 1242
Db 1142 AQ-----PCLSSGFNSPALWIKGSHMSLDNPDYQODEFPKREYNGIFKG-PT 1191
Qy 1243 AENPEYGLDVP 1254
Db 1192 AENAEYLKRVAP 1203

RESULT 4
Q9YH40
ID Q9YH40 PRELIMINARY; PRT; 1165 AA.
AC Q9YH40;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percomorphia; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
NCBI_Taxid=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; Pubmed=9582016;
RA Dmitriyevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AAD10500.2; -.
DR HSP; P11362; IFGK.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR001245; Gram_pos_kinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR Prodom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

SEQ SEQUENCE 1165 AA; 129614 MW; 7f7EB38D8771A74E CRC64;
 Query Match 39.2%; Score 2659; DB 13; Length 1165;
 Best Local Similarity 45.2%; Pred. No. 1e-192;
 Matches 575; Conservative 162; Mismatches 395; Indels 140; Gaps 29;

1 MELAALCRWGLLALPPG-AAST---QVCTGDMKRLPASEPTLDMKRLHYOCQV 55
 4 LELEL-----LILLLLSLIGRCSTDPDRKYCCQGSNQMTM---LDNHYLKKKKKYGSCNV 56
 56 VQGNLELTPTNASLFLDIOEVGVYLIANQVQVPLRIYRGOLFEDNYALA 115
 57 VLELELTPTQENODLSFLOSIOEVGGIVLIANNEVSTIPLVNLIRLQNLVEGNTLL 116
 116 VLDNCPPLNNTPTVPGSPGLRELQRLSTELIKGVLIQRPOLCYODTILMKDI FKH 175
 117 VMENYQK-NPSSP--DYVQYGLKQLQLSNLTIELISGVKVSHPNLLCNVEITNMWDIVDK 173
 176 NNOLALTLLDTNRSRACHPCSPMKSGSRGSESEDCQSLTRYCAGGC-ARCKGPLPY 234
 174 TSNPTNMLIPIHAERQCKCDPCGVNGSCWAPRGHCQKFTKYLCAEQCNRRCRGPRI 233
 235 IKANSKRTIGITELKHSCLCLHFNHSGICEHCPALVTYNTDFEESMPNEGRYTGAS 294
 234 CNEHCAGGCTGPRAIDCLCRPDNDGCTKOTCPPKITDIYSHQVNDENITKTTGAA 293
 295 CVTACPYNLSTDVGSCTLVCPPLHNOETVADGTQRCSEKSKPCARYCYGLMEHLRVR 354
 294 CVAREPSNYVTE-GACVRSKCSAGMLEVD-ENGKRSKPCDGVCPKYCDGIGISLNTI 351
 355 AVTSANIQEPAGCKKIFGSLAFLEPSFDGPASTAPLOPQLOVFETLEITGYLYISA 414
 352 AVNSTINGSFNCTKINGDIILNNSFEFGDPHYKIGIMDEHMLNLTVEITGYLYIMW 411
 415 WPSLPLDPLSFQNLQVIRGRILHNHAYS-LTQGLGSMWGLRSLRGLSLALIHNTH 473
 412 WPNMTSLSFQNLQVIRGRITTSRGSFVVQVSHQMLGSLKVSASANTLTKTPQ 471
 474 LCEVHTVPMQDLFRNPQALHTANRPEDECEVGEGLACHOLCARGHSGMGPPTQVNCQ 533
 472 LRYASTINMRRLFRSEDSQIEYDART-----ENQCNNECSDDGCMGPFIMCVSLH 524
 534 FLRGQEVCECRYLQCLPREVYNAHNLPCHPGCPQNGSVYTCGPRAQCVACAHYKDP 593
 525 VDRGSCVASCNLLQGEPRBAQVDRGCVCHOELVQDLSITCYGPGPANCSCAKHPQDG 584
 594 PCFVACPSGVKDPDLSMPLMKPPEDEGACOPGICNTHSCVDLDDGCAEQASPLTS 653
 585 POCIPRCPHGLGDDTL-IMKTADKMGQCOPCQNCCTQCGSGFSGCGD-IVSHSL 642
 654 IVSAVGLLVVGVFGLIKRQOKIRKYTRRLLOETELVEPLTPSGAMPNQOMR 713
 643 AVGLVGLTLTIVALLIVLRRRLIK-RKRTIRRLQKELELVEPLTPSGQAPNQAFR 701
 714 ILKETELRKVKVLGSGAFGVYKGIWPDGENVKIPIAIVLRENTSPKANKETLDEAYV 773
 702 ILKETEFKXRVLGSGAFGVYKGLMNPDENIRIPIAIVLRENTSPKANKETLDEAYV 761
 774 MAGVSPVYSRLIGICTSTVOLVTOIMPGCLLDHRENRGLSGDILNMCQIAKGM 833
 762 MASVDHPHVRCLGICLTSVOLTQIMPGCLLDYRQHOERICGMWLNMCQIAKGM 821
 834 SYLEDVRLVHRLAARNVLVKSPNHVKTDFGLARLLDIDETEHADGKVPKIMMLES 893
 822 NYLEERHLVHRLAARNVLLKPNHVKITDFGLSKLLTAEBKEVQAHGKVPKIMMLES 881
 894 ILRRRTTHOSDVWSYGVTTWELMTFGAKPYDGIPIAREIPLLEKGERLPOPICTTIDVM 953
 882 ILQWITTHOSDVWSYGVTTWELMTFGSKPYDGIPIAKIASVLENGERLPOPICTTIEVM 941
 954 IMVKCMNIDSECGPRPRELVSFSSRMARDQRFVJONEDLGASPLDSTFYSLIEDD 1013
 942 IILKCMNIDSSRPREFELVEFSQMAKDPSTRYLVQ---NLSPSPDRRLFRSLSSDD 998

1014 MGLVDABEYLVPQGGFCPPDPAPAGAMVHRRSSSTRSGDGLTLGLEPSEBAPRS 1073
 999 --DVYDADEYTL-----RKYRIN-RQGS-----E 1019
 1074 PLASEGAGSDVFDGLMGMAKGLQSLPTHPBPLOYSDPTV-PLPSTDGVAFLT 1132
 1020 PCIPPNCH-----PVENSIALRYISDPTQMALEKLDGH----- 1054
 1133 CSPQREYVNPDPVAPQ-----PSPRE-----GLP-AARPGATLERAKTSLPGKNG 1179
 1055 -----EYVNOGSESTRSLSDIYNPNYEDLTDGKGPVSLSQEAEATNFSREYLVNTONS 1109
 1180 VVKDVFAGAVENPEYLTTPQGAAPQPHPRPAPSPFDNLVYWDOPPERGAPRSTFGK 1239
 1110 L---PLVSSGSMDDPV---QAG-----YQAAV-----LPQTGALTGNGMF 1144
 1240 TPTAENPEYIGL 1251
 1145 LPAENLELYLGL 1156

RESULT 5
 Q9W6F6 PRELIMINARY; PRT; 1137 AA.
 ID Q9W6F6
 AC Q9W6F6
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERBB4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDBRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nr1) and erbB4 transcripts in
 RT embryonic chick hindbrain.";
 RL Mol. Cell. Neurosci. 13:237-258 (1999).
 DR EMBL: AF121963; AAD31764.1; --
 DR HSP: P11362; IFCG.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001368; TNFR_c6.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 1.
 DR Pfam: PF02757; YLP_2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYKc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KM Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 38.8%; Score 2635.5; DB 13; Length 1137;
 Best Local Similarity 46.5%; Pred. No. 5.9e-191;
 Matches 530; Conservative 167; Mismatches 365; Indels 77; Gaps 24;

161 LCYQDTILMKDIFHKNNQALTLITNRSRACHPCSPMKSGRSGSESDCQSLTRTV 220

```
DB 3 LCPADITHMODIVNRPMASNFTLVPTNGSSGGRCHKSCGTG-RCMGPTENHCQTLTKTVAC 61
QY 221 AGGC-ARGKGPLPOYIKANSKFIGITELKSHDCLALCPHNSIGTELHCPALVTYNTDTF 279
DB 62 AEQCDGRGCVGVYSDCCRRGAGGSGGPRDIDCFACNMFNDSGACVTCQPPQFVFNPTTF 121
QY 280 ESMNPEGRYFEGASCVTACPNYUJSTDVSGCTVLCPLHNOEVTAEQDRECKSKPCA 339
DB 122 QLENNHNAKTYGAFVCKKCHNFV-VDSSSCVACBSSKHEV-BENGIKMKCPETDICI 179
QY 340 RVCYGLGMEHLREVAVTSANIIOEPAGCKIFGSLAPLPESPDGPASNTAPLOEOLQV 399
DB 180 KACQIGTGLSVSAQTVDSNIDKFINCTKINGNLIFLVTHGHPYHTIAINEXKINI 239
QY 400 FETLEITGCVLYISAMPDLSLDLSFQULQVIRGILNNGAVSLTLOGLISMLRSLR 459
DB 240 FQVAREITGLYNISWPNEMTDFRVFSNLVTIGGALVSGSLLLKQOGITSLQFOSIK 299
QY 460 ELGSGALALHNNHNLCPVHTVPMQDLPNNPHQALHTANRPDECEVGEGACLOLCARGH 519
DB 300 QISAGNIYITDNSNLCTHYHTNMSTLFTSPQKTYIHNNKKAENCTADGMCNCLSSDG 359
QY 520 CMGPGPTQCVNCSQFLRGQECVEBCRVLQGLPREVNAHCLPCHPEQCP-ONGSVTCFG 578
DB 360 CMGPGPDQCLSKCRFIRGRTCIESCNLVGEFREFPANGSVCEQDPOCEKMDNMITCYG 419
QY 579 PEAOQVACAHYKORPFVACRPSGVKRDLSYMPIMKRPDEGAOQPCRPINTHSCVDLD 638
DB 420 GPHRCHTKCFHKGQPNCEKPRDLOQANSF--FKYADEBRECHPHCPNCTOCCRPA 477
QY 639 DKGC-----PAEQRASPLTSIVSAVY-GILLVVLGVNFGILIRKQOKIRKYT 686
DB 478 SHDCIYPTWQSTLPGHAR--TPL--IAAGVIGLPIIVIMKLTAAVYVRKRSIK-KKRA 533
QY 687 MRLLQETELVEPLTPSGAMPNOQMRILKTELKRVKVLGSGAGFYVYKGIWPDGENV 746
DB 534 LRRLF-ETELVEPLTPSGTAPNOALRIKETELKRVKVLGSGAGFYVYKGIWVEGETV 592
QY 747 KIPVAIKYLRNTSPKANKETLDEAYVMAVGSPVVSLLGLCLTSTYQVLTQMLPYCCL 806
DB 593 KIPVAIKYLRNTSPKANKETLDEAYVMAVGSPVVSLLGLCLTSTYQVLTQMLPYCCL 806
QY 807 LDHVENRGRGLSGODLNMCMQIAKMSYLEEDVLVHNRDLAARNLVSPNNHKTITDGL 866
DB 653 LDVYHEHNDNIGSOLLNMCVQIAKGMVYLEBRRLVHNDLAAARNLVSPNNHKTITDGL 712
QY 867 ARLLDIDETEHADGGKVPYKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 926
DB 713 ARLLGDEKEYNADGKMPYKMALECIHYRFTHQSDVMSYGVTVWELMTFGAKPYDGI 772
QY 927 PAREIPLDLEKGERLPORPCTIDVYMMVCMMDISCRPRERLVSEFSGMADPQRF 986
DB 773 PTREIPLDLEKGERLPORPCTIDVYMMVCMMDISCRPRERLVSEFSGMADPQRF 986
QY 987 VVIQNEI-LGPASPLDSTFYSLLEDDMDGLVDAEYLVPOQGFCDPAPAGAGMWH 1045
DB 833 LVIGGDDMKLPSPNDKFPQNLDEEDLEDDMDAEYLV-QAANIPIPIYTSSTRIDS 891
QY 1046 RHRSSTRSGGADLTLLGLEPSEBEAPRS--PLAP-SEGASDVFDGLGMAKAGLOSLP 1102
DB 892 NRNOFVYRDGGYAAEQGV-PMPYRAPGCIIEAPAPQAQATAIEFDTCNGTLRQVATL 950
QY 1103 THDSPLORYSEDPVPLPS-----ETDGVAPLTSQPEVYNQGDVAPQPSRPEG 1155
DB 951 AKESSTORYSADPLVFIPEVYIRGELDEDDGIMTMRDKPKTDYINPVENPEVSRKNG 1010
QY 1156 PLRPA-REPAATLERAKTLSPGKNGVVDV-----AFGAVENPEYITPOGAPQ 1206
DB 1011 DLQAVDNPEYHN-----APNGQPKADEVYNEPLVYNTFANTLENLEYL-----K 1055
QY 1207 PHRRPAPAPADNLVYWDPPPERGA--PSTFKCTPT-----AENPEYL 1249
```

```
DB 1056 NNLEPEKAKFAPDNPYNNHSLPPRSTLQHPDVLQESYTKFYKQNGRIRPIVAENPEYL 1114
RESULT 6
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMBLrel. 03. Created)
DT 01-MAY-1997 (TREMBLrel. 03. Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21. Last annotation update)
DE Etb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; Pubmed=10077531;
RA Gellner K., Brenner S.,
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGRF_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A335039258B647E9 CRC64;
Query Match 33.0%; Score 2241; DB 13; Length 1328;
Best Local Similarity 39.5%; Pred. No. 6-6e-161;
Matches 515; Conservative 153; Mismatches 411; Indels 224; Gaps 32;
QY 9 WGLLALPLP--GAASD-----VCTGDMKRLRPASRPTHMDMLHLYOGCVOVGNTLE 62
DB 4 WRLIMCVASRLRAASSQTOEAVCPGTONGLSSTGSOENYLNKDRYKGCETIIMGNDLEI 63
QY 63 TYLPFNASLFLQDIOEVQGVVLTAAHNGVQVPLQRLAIVRGTOVFEDNVALAVLDNDP 122
DB 64 TQIESNDPFLKTIREVTVGLVLAHNFQEIPLQQLVIRKNSLYERRFALSVFLN-- 120
QY 123 LNNTPVYVAGSPGGLREIQLSLTEILKGVLYIQNPOLCYODTLIMKDIFFKNNQLALT 182
DB 121 ----YKDG--PSGLNQLGMLNLTBIIDGVIQIINNKLRLGPRWYMHDIIR-NNDAPIE 173
QY 183 LIDNRSAPCHPCSPMKGRKSCWSESSDQSLTRTVAGGC----- 224
DB 174 IOFNGERGCH--KSC-GNYCWMGKQDCOILKTVCAPCQNDRCFGTSPRDCCHIECA 229
QY 225 ARCGRLPOYIKANSKFIGITELKSHDCLALCPHNSIGTELHCPALVTYNTDTFESRPN 284
DB 230 AGCGKPL-----DTDFCARLRFNDSACVPOCCQTLTYNKQTFOMETN 272
QY 285 PEGRYTFGASCVTACPNYUJSTDVSGCTVLCPLHNOEVTAEQDRECKSKPCA 343
DB 273 PNAKYQVGSICVSCQPTHFV-VDSSSCVSVPRDMEV--ERSGRQCELSGCLCPKVC 329
QY 344 GLGMEHLREVAVTSANIIOEPAGCKIFGSLAPLPESPDGPASNTAPLOEOLQVETL 403
DB 330 GTGAE---QROTVDSSNIDSFINCTKIQSGILHFLVTVTGILGDDFQVNPVLDAAKLEVFRTV 386
```



```

Qy 404 EETGTYLISAWPDSLDLVFONLQVIRGRILNNGAVSLTLOGLISGLRLSRELS 463
Db 387 REITDILNISQWPKELNDLVSVSLTITQGRSLFRFSLMWIRIPTLSLRLSREISD 446
Qy 464 GLALIHNTLHCFVHTVMDQLFRRNPH-QALLHTANREDCEVGBGLACHOLCARHGCMG 522
Db 447 GSVISQNAHLCHHTVMTOLFGRSRRVANSLSNRMAECVADGRVCDELCSGSGCMG 506
Qy 523 PGPTQVNCQSLRGQCEVCEVRLVQLPREVYNAH-CLRCHPECOPONGSVTCFGEA 581
Db 507 PGPOCLSCRNSYSHGTCVACGHFNSGIPREPAGLNGVCAHCECKQOTGKASCTGCA 566
Qy 582 DQVACAHYDPRPCVARGCVKPRDLSYPMIKPRPDEGACORPINCISCVDLDDXG 641
Db 567 DECAACRFRDGPYCMSSCPAGVN-DEGKGLIFKPRNKGCEPHONCTGCGSGPGIND 625
Qy 642 CPABORASPLTISAVVGLLVVVLGVV-----GILKRBQOKIRKXTMRLLOET 694
Db 626 C---LEAARLTSSGOITGLAGVPAGLIFCLVLEFLGMLYHRLALIRKKAARRRYLSEG 682
Qy 695 ELVEPLTPSGAMPNOQRILKETELRKVYLGSGAFGTYYKGIWIPGENVKI PVALKV 754
Db 683 ESFEPLGP-GEKGTFRKHARILKPSDLRKIKPLGSGVFGTYSKGFWIPGETVKIPVALKT 741
Qy 755 LRENTSRKANKEILDEAVTMAGVSPYSLIGTCLSTVOLVQMLMYGCLLDHVENR 814
Db 742 IQDSGGOTFEITDHLISWGLSHPIYVRLIGICPGCLDLVYQLSHSGSLLEIRHOK 801
Qy 815 GRLLSQDILNMCQIYAKGMYLEDVRLVHRLDAAARNVYKSPNHYKIDFGALRLDDE 874
Db 802 TSLDPQLNMCVQIYAKMYLLEHRYVKNLAAARNILKNDYQVQISDYVADLLVPDD 861
Qy 875 TEYHADGKVPYIKWMALESILRRFTHQSDVMSYGVTVWELMTGAKPYDIPAREIPDL 934
Db 862 KKYYVSEKTPYIKWMALESILFRYTHQSDVMSYGVTVWEMSGAEFYVASVQGEVPSV 921
Qy 935 LEKERIPORPITCTIDYVMYKCMYIDSECRPRFREIVSFSMARDPQRFVYIQNEDL 994
Db 922 LEKERISOPACITIDYVMWVKCMYIDENRPTFKELASDFTMARDPREYVILRMG- 980
Qy 995 GPASPLDSTFYRSLDEDDMDLVDAEYLVYQOQFCPRDPAFGAGVHHRHSSSRSS 1054
Db 981 -----EDSGKREF-----RGSER- 995
Qy 1055 GGGDLTLGLEPSEEEAPRSLAPSEAGSDVFDGLMG--AAKGLSLPHTHPSPLQ- 1110
Db 996 --GILBADLEDEE-----GLGDRFATPISLOPSSWETSPOI 1032
Qy 1111 -----RYSEDPYVPLPSETDGYVAPLTCSPQ- EYVNO-----PDVR 1146
Db 1033 NSYWMVMTOLRYD-----FAVSOGGHIGYLFMSPSPVDITIRQLWYORSRLSSVRLPDRS 1086
Qy 1147 POPPSPEGPL--PAARPAATLEAKTILSPCKNGVAVKDVAFGAVENPEYVLPPOGAA 1204
Db 1087 AFRSSSEAEELCEGAQACAGIFRYR-----FSSERGN-----POG- 1122
Qy 1205 POPHPAPAFSAPFADNLVYWDODPERGAPSTFKGTPTAENPE 1247
Db 1123 -----OQRKLTSTASSPSSFKTMADEDE 1146

```

```

OC Anopheles.
OX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN: SUA.
RA Lyceet G.3.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ13655; CAC35008.1; --
DR HSSP: P11362; 1GCK
DR InterPro: IPR000345; CyC_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; EGF_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fu_7.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR Receptor.
FT NON TER
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D86967724F07 CRC64;

Query Match 28.9%; Score 1961.5; DB 5; Length 1433;
Best Local Similarity 32.1%; Pred. No. 1,2e-139;
Matches 465; Conservative 196; Mismatches 401; Indels 385; Gaps 36;

Qy 26 CTGTDMKRLRLPASETHLDMRLHLYOGCOVQGNLETLVLPNALSFLQDIOEYQYVL 85
Db 1 CIGTNGMSVAPANEHYKYLRLDRYNTCTYDGNLEITWIONITDNLQIRETVGYL 60
Qy 86 IAHNOVQVPLQRLIRVGTQLF-----EDNYALAVLDNGPPLNNTTPVTGASPGGLREL 140
Db 61 ISLYDLPOVILPRQIIRGRTTFKLNKEEAYGLFV-----SFSHMTL 104
Qy 141 QLRSTELILKGVLIQNPOLCYODTILMKOI-FHKNNQALLTLIDTRSPACHGSPMC 199
Db 105 ELPLRLDILGSAVFENNYNLCHMKSLNWEIILLAPOTSQWYTFNFFSSPERVCPCHPSC 164
Qy 200 KGSRCWGESSEDCSLTRTVACAGCA--RCKGPLPOYIKANSKFIGITELKHSDCLACTH 257
Db 165 EVG-CMGEAGNACGRFSKLNCSPOCSGRCGPRPRECHLFCAGGCTGPTQSDCLACKN 223
Qy 258 FNHSGICELHCPALVTYNTDTFESMPNPRBGRYTFGASCVTACPYNYLSTDVSGCTLVPL 317
Db 224 FYDGVCKQECPRMQIYNPTNYFEPBPDPKAYAGATCVRKCP-EHLKXGACVYRKCPK 282
Qy 318 HNGVETADGTQREKSKPCARVCTGICGHEHLREVAVTSANIQEFACCKKIFGSLAF 377
Db 283 GKMPQNSE-----CVPKGVCPKTCPEBGIVH-----SDNIGNYDCDTIIGSLEIL 329
Qy 378 PESFDPGPASNT-----APLOPEOLOVETLEITGYVYISAWPDSLDLVFONLQ 429
Db 330 DQSFDFGOVYVYTNSSFPSPRYIKIPDLLEVFSTYKEITGFINTIQAHPNFTTLNYPFRLE 389
Qy 430 VIRGRIILHNGAY-SLTLOGLISWLGRLSRELSSGLALIHNTLHCFVHTVMDQLFRN 488
Db 390 VVGGROQKEMLFASVYIVKTSLSKLELSKRVVSGSIVILENSDLCFVEDIDMSEIKKS 449
Qy 489 PHQALLHTANRPEPECVGBGLACHOLCARHGCMGPGPTQVNCQSLRGQCEVCEVRLQ 548
Db 450 SDHEWVQKNRMATCHEBGMCSBQCSKAGCMGKGEQCLCKNVYKGCICLDSCK--- 506

```

```

QY 549 GLPREY-VNARHCLPCHPECOPOONGSVTCFGEADQVACAHYDPFPCVAPRC----- 601
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 SLPLVSVDSKTCDDCHQEKD-----FCVGNEDNGSCMNVNDRGRCVAECPTTKAM 561
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 602 -----SGVKPDLSTWPKFPD----- 618
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 562 NGTCINCHKTGCGRPRDTIAPDGCISCDAIIGSDAKIERCLMKOESCPDGYSDVYL 621
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 619 -EEG----- 621
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 622 QEEGRLKOLSGKAVCRKCHPRCKCTGTFGHEFCOECTGYKKGBQCEDECPDYANEE 681
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 622 --ACQCPRICT-----HSCVDL-----DD-----KGCRAEO----- 646
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 682 TRICLPCHQECRGCHGLGDHHECRNLKLFEGDPDYDNTPTTCVSNCPASHPYKRPQEA 741
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 647 -----RASPLTISVAVGILLVVLGVNGI---LTKRQOKIRKRYTM 687
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 742 GKIGPYCSADSMOSGLRIEPTQYKIVMSVMAIILCVGVGIAFVLFSRHKNKKDAVKM 801
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 688 RRLIOTELVEPLTPSGAMPNOAQMRILKETELRKVKVIGSGAGTVYKGIWIPDGENVK 747
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 802 TMLAGCEDESPRLPSVNGPRLTKLRIRKEAIRGCVLGMGAFGRVYKGVMMPEGESVK 861
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 748 IPVAIKVLENTSPRANKELIDEAYVMAGSPYVSRLGICTSTVOLVTQMLPYGCLL 807
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 862 IPVAIKVLENTSPRANKELIDEAYVMAGSPYVSRLGICTSTVOLVTQMLPYGCLL 921
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 808 DHVENGRGLASODLLKWCQIAGKMSYLEVRLVHRLLARNLVYSPNNVKTITDELA 867
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 922 DYVNNNDKIGSKALLNMSTQIARGMAVLEERRLVHRLLARNLVYSPNNVKTITDELA 981
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 868 RLIDIDETEHYADGKVKIKMMALLESILRRFTHOSDVYSGVTVWELMTGKAPYDGP 927
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 982 KLIDFDEDEYAAAGKPKIKKALECIRHVRFTSKSDVMAGITITWELTTGAPRYENVP 1041
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 928 AREIPDLLEKGERLPORPCTIDVYMTWVKCMIDSECRPRFRELVSFSSMARDPORFV 987
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1042 AKDVPREIIEGKLPQPDICSLDYVCILSCVILADARPTFKQLAETFAEKARDPGRYL 1101
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 968 VIONEDIGRASPRLSTFRSLLEDDMDGLV----- 1018
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1102 MI-----PGKFKMLRPSYTNODEKDLRLTAPVMAAAAAAGASNVDPSTIA 1152
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1019 DAEELVPOOGFPCPDPAFGAGMNNHRRSSSTRSGGDLTLGLEPSEEPARS----- 1073
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1153 ETDXYLQPKTRPSIMLRPSA-----VEPS-DEMPKSLRYCK 1168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1074 -PLAP---SEGAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYVA 1129
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1189 DPLKPDDETIDGHEKVE-----GVGQIR-----LNLPLDEDDVLM 1222
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1130 PLTSGPQEVYVNDVVRQPPSPREGPLPAPRAPATLIERAKTISPGKNGVYKVFARGG 1189
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1223 P-TCOSQ---NOS-----TPG---YMLDIGVPA 1243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1190 AVENPEYL-----TPGGAAPQPPAPPAFPANLYYMDODPERGAPSTFKGT 1240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1244 SVNDPEYLMGSTOAIAGLAQSGM--PHTPP-----PNTTNGM 1280
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1241 PTANPE 1247
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1281 PTHQHSQ 1287
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK72;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
Herstein..

```

```

GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF17761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

```

```

Query Match 26.0%; Score 1766.5; DB 4; Length 419;
Best Local Similarity 86.3%; Pred. No. 1.2e-125;
Matches 340; Conservative 5; Mismatches 42; Indels 7; Gaps 2;

```

```

QY 1 MELALCRWGLLALLPPGAASVQCTGTDKRLPASPEHLLMLRLHYGCGVQVGNL 60
Db 1 MELALCRWGLLALLPPGAASVQCTGTDKRLPASPEHLLMLRLHYGCGVQVGNL 60
QY 61 ELTYLPNASTSFQDIQEVGVYLIANNQVQVPLQRLIVRGTOU.FEDNYALAVDNG 120
Db 61 ELTYLPNASTSFQDIQEVGVYLIANNQVQVPLQRLIVRGTOU.FEDNYALAVDNG 120
QY 121 DPLNNTPTVTCASPGIARELQRLSLTEILKGVLIQRPOLCYODTILMKQIFHNQOLA 180
Db 121 DPLNNTPTVTCASPGIARELQRLSLTEILKGVLIQRPOLCYODTILMKQIFHNQOLA 180
QY 121 DPLNNTPTVTCASPGIARELQRLSLTEILKGVLIQRPOLCYODTILMKQIFHNQOLA 180
Db 121 DPLNNTPTVTCASPGIARELQRLSLTEILKGVLIQRPOLCYODTILMKQIFHNQOLA 180
QY 181 LTLIDITRSRACHPCSPKCSKRCWGESSEDCQSLTRTVACAGGACRCKPLPOYIKANSK 240
Db 181 LTLIDITRSRACHPCSPKCSKRCWGESSEDCQSLTRTVACAGGACRCKPLPOYIKANSK 240
QY 241 FIGITELKHSQCLACLPNHSIGICELCPALVTYNTDFESMPNPEGRYTGASCVTACP 300
Db 241 AAGCTGPHSDCLACLPNHSIGICELCPALVTYNTDFESMPNPEGRYTGASCVTACP 300
QY 301 YNYLSTVGSCITLVCPHNOEVTAEADGTQRCCKSKPCARVCYGIEMHLREVAVTSAN 360
Db 301 YNYLSTVGSCITLVCPHNOEVTAEADGTQRCCKSKPCARVCYGIEMHLREVAVTSAN 360
QY 361 IQEFAGCKKIFGSLAFLPESFDGPASNTAPLQ 394
Db 356 LRMQPG--PAHPVLSFLRPSWDLVSASFYSPLAP 367

```

```

RESULT 9
Q9R2X1
ID Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

```

DR EMBL, J0027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;
Query Match 25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 1.2e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;
QY 889 MALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICT 948
DB 1 MALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICT 60
QY 949 IDVYIMVWKCMIMDSECRPFRELVSFSRMARDPQRFVVIQNEIDLCPASPLDSTFYRSL 1008
DB 61 IDVYIMVWKCMIMDSECRPFRELVSFSRMARDPQRFVVIQNEIDLCPASPLDSTFYRSL 120
QY 1009 LEEDDDMDGLVDABEYLVPOOGFFCPDPAFAGGVVHHRSSSTRSGGGLTGLGPESEE 1068
DB 121 LEEDDDMDGLVDABEYLVPOOGFFSPDPLTGSTAHRRHSSSARSGGGLTGLGPESEE 180
QY 1069 EABRSPLAPSEGASDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYV 1128
DB 181 EPRRSPPLAPSEGASDVFDGDLAVGVTKGLQSLSPHLSPLQRYSEDPITLPLPETDGYV 240
QY 1129 APTTCSPQPEYVNPQDVRPQPPSEBEGPLPAARPAGATLERAKTSLFGKGVVQVAFAG 1188
DB 241 APLACSPQPEYVNPQDVRPQPPSEBEGPLPAARPAGATLERAKTSLFGKGVVQVAFAG 300
QY 1189 GAVENPEYLTPOGGAAPQHPAPAFSPFNLYVWDDPREKAPSTFEGTPTAENPEY 1248
DB 301 GAVENPEYLTPOGGAAPQHPAPAFSPFNLYVWDDPREKAPSTFEGTPTAENPEY 360
QY 1249 LGLDVPEV 1255
DB 361 LGLDVPEV 367
RESULT 10
ID 086712 PRELIMINARY; PRT; 729 AA.
AC 086712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxId=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.,
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PFO0069; pkinase; 1.
DR Pfam; PFO2813; Retro_M; 1.
DR ProDom; PDOM0001; Euk_pkinase; 1.
DR SMART; SMO0219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00113; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84DF6914EFD63 CRC64;

Query Match 25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 9.5e-122;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;
QY 569 PQNGSVTCFPEPADOCVACAHYKDPFCVACRCGSGVVPDLSTYMIWKFPPEEGACQPCPI 628
DB 141 PEETAAPKTP--DHCKKCHAFIDGPHCVACPAIGENDTL-WKRYAANAACOLCHP 197
QY 629 NCTHSCVVDLDKCGCPAORASPLTSIYSAV-GILVAVVGVGVGILIKRQKIRRYTM 687
DB 198 NCTRCGCGPLBEGP--NSKTPSTIAGVVGILCLVAVGLGILYLR--HIVRKTLL 253
QY 688 RLLQETELVEPLTPSGAMPNOAQMRLKETELRKVKVLSGAFVYKGIWIPDGENVK 747
DB 254 RLLQERELVEPLTPSGEAPNOAHLRLKETEFKXVAVLVSAGAFVYKGIWIPDGENVK 313
QY 748 IPAVAILKRENTSPKANKELIDEAIVAGVSPVSLGICLTSTVOLVTLMPYCCLL 807
DB 314 IPAVAILKRENTSPKANKELIDEAIVAGVSPVSLGICLTSTVOLVTLMPYCCLL 373
QY 808 DHVRENRGRGSDLLNWCMIKMSYLEDVLRHDLAARVLYKSPHHVKTDFGLA 867
DB 374 DYTREHKNIGSOYLNMVQIAKGMVLEERRLVHRDLAARVLYKTPQHVKITDFGLA 433
QY 868 RLIDIDETEHADGKVPIMMALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI 927
DB 434 KLGLADBEKEXHAAGKVPIMMALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGI 493
QY 928 AREIPDLLEKGERLPQPICTIDVYIMVWKCMIMDSECRPFRELVSFSRMARDPQRFV 987
DB 494 ASEISSVLEKGERLPQPICTIDVYIMVWKCMIMDSECRPFRELVSFSRMARDPQRFV 553
QY 988 VIO-NEIDLGPASPLDSTFYRSLLEDDMDGLVPAEYLVPOOGFFCDDPAPAGGVVHHR 1046
DB 554 VIQGDERRMLPSPDTSKFRYTLMEEDMEDIADADELVPHQGF----- 598
QY 1047 HRSSSTRSGGGLTGLGPESEEAPRSP-----APSEGASDVFDGDLGMAKGLQSL 1101
DB 599 -NSPT-----SKTPLLSSLSATSNNSATNCID-----RNCQGH 631
QY 1102 PTHDPSPLORYSEDPTVPLPSET--DGYAPLACSPQPEYVNPQDVRPQPPSEBEGPLA 1159
DB 632 PVVEDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVNG--LMPKKS----- 675
QY 1160 ARPAGATLERAKTSLFGKGVVQVDF-----AFCGAVENPEYL 1197
DB 676 -----TAVVQNOIYNNISLTAISKLPMSRYQNSHTAVDPEYL 715
RESULT 11
ID 086714 PRELIMINARY; PRT; 567 AA.
AC 086714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxId=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.,
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.

```

DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase.1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase.1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; 6CD9CBA7ADF225E1 CRC64;

Query Match 25.3%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 9.3e-122;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQVACAHYDPPFCVACRCPGKPDLSYPIKWFPEDEAGCOPCPINCHSCVDL 637
DB 1 GP--DHCKCKAHFDGPHCVACACPAVLGENDTL-VWKYADANAVCOLCHPNCGRCKGP 57
QY 638 DDKCPAEQRASPLTSIVSAVY-GILLVVLGVVFGILLKRRQKIRKYMRRLLQETEL 696
DB 58 GLECCP--NGSKTPSIAAGVVGGLLVVGLGILGLRRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOQOMRILKETELRKVVLGSGAFGVYKGIWIPDGENVKIPIAIKYL 756
DB 114 VEPLTPSGEAPNQHILKETEFKVVVLGSGAFGVYKGLMIPGEGKVIPIAIKELR 173
QY 757 ENTPKANKELLDAYVAGVGSPIVSRLLGICTSTVQLVTLMPYCGCLLDHVENKGR 816
DB 174 EATSPKANKELLDAYVAGVGSPIVSRLLGICTSTVQLVTLMPYCGCLLDYIREHNDN 233
QY 817 LGSODLNMCMQIKGMSYLEVDVLRHDLAARVVLVKSPPHVKITPFGARLLDIDETE 876
DB 234 IGSQYILNMVCQIKGMNYLEERLVRHDLAARVVLVKTPOHVKITPFGAKLLGADKE 293
QY 877 YHADGKVPKIMALESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIDPLE 936
DB 294 YHADGKVPKIMALESILRRRFTHQSDVMSYGVTVLMTFGSKPYDGIIPAREIDPLE 936
QY 937 KGEPLPQPICTIVYIMVYKCMVIDSECRPRRELVSFSSMARDPORFVITO-NEDLG 995
DB 354 KGEPLPQPICTIVYIMVYKCMVIDSECRPRRELVSFSSMARDPORFVITO-NEDLG 995
QY 996 PASPLDSTFYRSLDDDMGDLVDAEYLVPOGFCFCDPAPAGAGVNHHRHSSSTRSG 1055
DB 414 LPSTTDSKFTKTLMEEDMEDIVDAEYLVPHQGF-----NSPST-- 454
QY 1056 GGDLLTLGLEPSEEPSPPL-----APSEAGSDVFDGLMGAKGLQSLPTHDPSPLQ 1110
DB 455 -----SRPLSLSLATSNNSATNCID-----RNGQGHPRVDEDSFVQ 491
QY 1111 RYSEDPVPLPSEI--DGYVAPLTCSPQPEYVNOVDVPPQPSREGPLPAARPAATLE 1168
DB 492 RYSSDPPLPSEIDGFL-----PAPEYVNO--LMPKSPS----- 526
QY 1169 RAKTLSPKNGVVDVF-----AFGAVENPEYL 1197
DB 527 -----TAMVQNOIYNNISLTAISKLPMSRYONSHSTAVDNEYL 566

```

RESULT 12

```

ID 08MYVO PRELIMINARY; PRT; 412 AA.
AC 08MYVO:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Placentalia; Primates; Catarrhini; Hominoidea; Homo.

```

```

OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF18349; AA5586.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; kinase.1.
DR Pfam; PF02757; YLP.2.
DR ProDom; PD000001; Euk_kinase.1.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; EF_HAND; UNKNOWN 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397F3F27D2BC CRC64;

```

```

Query Match 25.0%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 2.1e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

```

```

QY 889 MALESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIDPLEKGERLPQPICT 948
DB 1 MALESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIDPLEKGERLPQPICT 60
QY 949 IDVYIMVYKCMVIDSECRPRRELVSFSSMARDPORFVITO-NEDLG 1008
DB 61 IDVYIMVYKCMVIDSECRPRRELVSFSSMARDPORFVITO-NEDLG 120
QY 1009 LEDDDMDLVDABEYLVPOGFCFCDPAPAGAGVNHHRHSSSTRSGGDLTLGLEPSEE 1068
DB 121 LEDDDMDLVDABEYLVPOGFCFCDPAPAGAGVNHHRHSSSTRSGGDLTLGLEPSEE 180
QY 1069 EAPRSLPAPSGAGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGV 1128
DB 181 EAPRSLPAPSGAGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGV 240
QY 1129 APLTCSQPEYVNOVDVPPQPSREGPLPAARPAATLEAKTLSPKNGVVDVFAFG 1168
DB 241 APLTCSQPEYVNOVDVPPQPSREGPLPAARPAATLEAKTLSPKNGVVDVFAFG 300
QY 1189 GAVENPEYLVPOGGAAPOP-----HPPA--FSAPFDNL 1220
DB 301 GAVENPEYLVPOGGAALSPFTLLPSAQRSTTSTGTTHQSGHAPRASKGLMQRQST 360
QY 1221 YYMD-QDPPER-----GAPSTFKCTPTAEN 1245
DB 361 WWTTCQCEPREGVARRSPDVSSGREGLTSAQIKRWGEPPTTSRGTCAR 410

```

RESULT 13

```

ID 064895 PRELIMINARY; PRT; 962 AA.
AC 064895;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Gag-v-erb-A-v-erb-B protein.
GN GAG-v-erb-A-v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OC NCBI_Taxid=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; Pubmed=1969616;
RA Bruskini A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
transforming potential of the oncogene v-erb-B."

```

```
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: X52209; CA36459.1; -.
DR EMBL: X52211; CA36459.1; JOINED.
DR HSSP: P10828; 2NUL.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001723; Sterm_receptor.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR001628; Znf_Casteroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00069; kinase; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHOMONER.
DR PRINTS: PR00047; STROIDINGER.
DR ProDom: PD000001; Euk_kinase; 1.
DR ProDom: PD000035; Znf_Casteroid; 1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00219; Tyrc; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
KM Transcription regulation; Transferase; Tyrosine-protein kinase;
SO ZINC-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791EAE95CE CRC64;

Query Match 24.4%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 1.6e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

Qy 541 VEEGRVAGLPRE-VVNAH-HCLP-----CHPCQ 568
Db 354 IEKQESYLLAFEHYINRKNIIPFWSKLMKVALDLMIAGAHASRLMKVECPTELS 413
Qy 569 PONGSVTCFGEADQCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDEGACQCPPI 628
Db 414 PQE-----VGP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VMKADANAVCOLCHP 465
Qy 629 NCHSCVDLDKCPRAEQRASPLTSIYSAV-GILLVVVLGVFGILIKRQOKIRKXTM 687
Db 466 NCTGCGKPGLEGCP---NGSKTPSIAGAVGGLCLVVGIGLGLYLRH-HIVRKRTL 521
Qy 688 RRLQERELVEPLTPSGAMPQAOHRLIKETELRKVVLGSGARCTVYKGIWIDGENVK 747
Db 522 RRLQERELVEPLTPSGAMPQAOHRLIKETELRKVVLGSGARCTVYKGIWIDGENVK 581
Qy 748 IPVALIKLRENTSPKANKEILDEAYVMAVGSPYVSRLGICLTSTVQLTQLMPIYGL 807
Db 582 IPVALIKLRENTSPKANKEILDEAYVMAVSNPHVCRLLGICLTSTVQLTQLMPIYGL 641
Qy 808 DHVRENKRGISQDLNWKQIAKMSYLEDLVLRDLAARNVLYKSPNHVKTIDFGIA 867
Db 642 DYIREHNDIGSQYLWMCVQIAKMSYLEDLVLRDLAARNVLYKSPNHVKTIDFGIA 701
Qy 868 RLDDIDETVYADGKVPIMKMALESILARRFTQSDVMSYGVYVWELMTGKAPYDGP 927
Db 702 KQLGADKEKYHABGKVPIMKMALESILARRFTQSDVMSYGVYVWELMTGKAPYDGP 761
Qy 928 AREIPDLKGERLPPPICTIDVYMIWKCMIIDSECRPFRELVSFSSMARDDPOFV 987
Db 762 ASEISSVLEKGERLPPICTIDVYMIWKCMSGADSRPFRELIAFSSMADDPFVYL 821
Qy 988 VIQ-NEDLGPASPLDSTFYRSLEDDMGDLVDAEEYIVPQGFPCPDPAAGAMVHR 1046
Db 822 VIQDEDMHLPSPYDSKYFTLMEEDMEDIVDADEYIVPHQGF----- 866
Qy 1047 HRSSSTSGGSDLTLGLEPSEEAAPRSPLASBEGAGSVFPODGLMGAKGLQSLPTTDP 1106
Db 867 -NSPST-----SRTPLLSLSATSN-----NSATKCIDRNGGH- 898
```

```
Qy 1107 SPLQRYSEDPVPLPSETDGYVAPLTCSPOEYVNOQDPVRPPSPREGPLPAARPACAT 1166
Db 899 -----PVREDGL-----PAREYVNO--LMRKXSTANVQNIYVIGLT 936
Qy 1167 -LERAKTLSPGKNGVNDVFAFGAVENPEYL 1197
Db 937 AISKLPMDSRYN-----SHSTAVNDPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID Q85468;
AC Q85468;
DT 01-NOV-1996 (TREMBLrel. 01. Created)
DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20. Last annotation update)
DE Avian erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCHI_Taxid=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scottling P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL: X06943; CA30024.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CA078AP4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 3e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDEGACQCPPI NCHSCVDL 637
Db 1 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VMKADANAVCOLCHPCRTGCKGP 57
Qy 638 DDKCPRAEQRASPLTSIYSAV-GILLVVVLGVFGILIKRQOKIRKXTMRRLQEREL 696
Db 58 GLEGCP---NGSKTPSIAGAVGGLCLVVGIGLGLYLRH-HIVRKRTLRLQEREL 113
Qy 697 VEPLTPSGAMPQAOHRLIKETELRKVVLGSGAFGYKGIWIDGENVKIPVALIKLR 756
Db 114 VEPLTPSGAMPQAOHRLIKETELRKVVLGSGAFGYKGIWIDGENVKIPVALIKLR 173
Qy 757 ENTSPKANKEILDEAYVMAVGSPYVSRLGICLTSTVQLTQLMPIYGLLDHVRENKR 816
Db 174 EATSPKANKEILDEAYVMAVSNPHVCRLLGICLTSTVQLTQLMPIYGLLDYIREHND 233
Qy 817 IGSQDLNWKQIAKMSYLEDLVLRDLAARNVLYKSPNHVKTIDFGIARLDDIDETE 876
Db 234 IGSQDLNWKQIAKMSYLEDLVLRDLAARNVLYKSPNHVKTIDFGIARLDDIDETE 293
Qy 877 YHADGKVPIMKMALESILARRFTQSDVMSYGVYVWELMTGKAPYDGPAREIPDLLE 936
Db 294 YHADGKVPIMKMALESILARRFTQSDVMSYGVYVWELMTGKAPYDGPAREIPDLLE 353
Qy 937 KGERLPPPICTIDVYMIWKCMIIDSECRPFRELVSFSSMARDDPOFVIO-NEDLG 995
Db 354 KGERLPPPICTIDVYMIWKCMSGADSRPFRELIAFSSMADDPFVLIQDERMH 413
```

```

OY 996 PASPLDSTFYRLSLEDDMDGLVDAEYLVPQGFCDPPAGAGMWHRRSSSTRSG 1055
DB 414 LPSPTDSKFYRTLMEEDMEDIVDADEYLVPHQGF-----NSST--- 454
OY 1056 GGDULTGLEPSEEPKRSPL-----APSEGASDVFDGDLGKAAGKLGSLPTHTDPSPLQ 1110
DB 455 -----SRTPLSLSLATSNSTNCTNDRNG-----H----- 481
OY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNOPOVPSPREBPPLPAARAGAT-LER 1169
DB 482 -----PYRENGFL-----PAPEYVNO-LMPKKPSTAMVOIQIYNTSLTAISK 523
OY 1170 AKTLPSPGKGVKDVFAFGAVENEPEYL 1197
DB 524 LPMDSRYQN-----SHSTAVDNEPEYL 544

RESULT 15
OY 09WVF5 PRELIMINARY; PRT; 655 AA.
AC 09WVF5;
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-NOV-1999 (Tremblrel. 21, last annotation update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibhe N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/10J, 129/SVJ AND 129/SVEVYAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearlsall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai Y., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi T., Fukuda S., Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Saito T., Okazaki Y., Gojobori T., Bono H., Kaetawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flischmann W., Gaasterland T., Gissi C., King B., Kochiya H., Kuhl P., Lewis S., Matsumoto Y., Nikolaic I., Pesole G., Quackenbush J., Schriml L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderas R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Guernicich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Matchonni L., Mashima J., Mazzarelli J., Mombaerts P., Norcose P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P., Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -
DR EMBL; AF275365; AAG38047.1; -
DR EMBL; AF275365; AAG38047.1; JOINED.
DR EMBL; AF275365; AAG38047.1; JOINED.
DR EMBL; AK004944; BAB33688.1; -
DR EMBL; AK004883; BAB33641.1; -
DR EMBL; AK004911; BAB33662.1; -
DR MGI; 95294; Egfr.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.2%; Score 1441.5; DB 11; Length 655;
Best Local Similarity 42.9%; Pred. No. 1,1e-100;
Matches 274; Conservative 98; Mismatches 244; Indels 23; Gaps 7;

OY 11 LLLALLPPGAA--STVCTGTDMKRLRLPASBETHLDMRLHYOGGVQVGNLELYLPTN 68
DB 14 LLLTALCAAGALAEKVKVCGTSNRLTQGFEDHFLSLQRMYNCEVILGNLEITYVGRN 73
OY 69 ASLSFLDIOEVQGVLLIAHNOVQVPLQRLRIYRGVTLQFEDNVALVLNDGPLMNTTP 128
DB 74 YDLSPFLKTIQEVAGVLLALTVERIPLENQIIRGNALYENTVALAILSN----- 124
OY 129 VTGASPGGLRELOSLRLTEILKGVLLIQRNPOLCYODTILWMDI----FKKNOLATLI 184
DB 125 -YGNRTGLRRLPRNLOEILIGAVRFNNPILCMDDITQNRDLYQVNFMSMSDL--- 180
OY 185 DTNRSRACHPCSPMCKGSRGWESSEDCQSLRTVACAGCA-RCKGRLPYOYIKANSKEFG 243
DB 181 -QSHPSRCPCDPCPCPNCSGWGGEENCQKLTIKICAGQCSHRGSRGSPDCCHQCAAG 239
OY 244 ITELKHSGLACLPHNSGICELHCPALVYNTDTFESMPREGGTYTGASCVTACPNY 303
DB 240 CTGPRESDCVQKQFODEATKDTCPPLMLYNTPTTYQMDVNPBGYSFGATCVKCKPNRY 299
OY 304 LSTVGSCTVLCPLHNOEVTAEADGTORCEKSCPKARVCYGLGMBHLREVAVTSANIOE 363
DB 300 VVTHGSGCVRAAGPRYYEV-BEDGIRKCKKDDGPRKVCNKGIGISFPDTSINATNIK 358
OY 364 FAGCKIIFGSLAFPESEFGDPASNTAPLOEQLVFEETLEITGYLISAMPDLSPLS 423
DB 359 FKCYTAISGDHILPVPFKGDSFRTPLDPRELEILKTVKEITGFLIQAMPDWTDLH 418
OY 424 VFQNLQVIRGILNNGASLTLQGLISWLGIRSLRELSGLALIHNTHLCSFNTVPM 483
DB 419 AFENLEIRGRTHKQGFSLAVGINTISLGRSLKESIDSDVVISGRNLCYANTIMWK 478
OY 484 QLFENPHQALHTANRPEDECVGEGLACHOLCARGCHGPGPTOCVNCOSFLRGOECYEE 543
DB 479 KLFGPRNQKTIKNNRAEKDKANNVHVNPLCSSGCGPBRDVCQNSKRGCEYK 538
OY 539 CNILEGEPRFEVNSECTIOCHPECLPQMMNITTCGRGPDNCGICQAHYIDGPHCVKTPAG 598
DB 604 VKPPLSTYPIKPFDEBEACQCPRCPTNCHSCVVDLDDKCG 642
OY 599 IMGENNTL-VWKYADVANNVCHLCHANCTYGCAGPGLQGC 636

```

Search completed: July 22, 2003, 09:00:39
 Job time : 54.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds

(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-103-117-14

Sequence: 6647
1 MELALCRWGILLALPPGA.....TFKGTPTAENPEVIGLDPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_73: *
1: pirl: *
2: pirl2: *
3: pirl3: *
4: pirl4: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6708	98.0	1255	1	A24571 protein-tyrosine k
2	5940	86.8	1260	1	TVRTHV protein-tyrosine k
3	5921.5	86.5	1254	2	I48161 p-185 precursor - f
4	3139	45.8	1210	1	GQHUR epidermal growth f
5	3108	45.4	1210	2	A53183 epidermal growth f
6	3093.5	45.2	1223	1	TVCHLV epidermal growth f
7	2986.5	43.6	1308	2	A47253 epidermal growth f
8	2679	39.1	1166	1	S06142 protein-tyrosine k
9	2418.5	35.3	1342	2	UC4387 kinase-related tra
10	2334.5	34.1	1339	2	epidermal growth f
11	1766.5	25.8	698	1	TVFVLV protein-tyrosine k
12	1703	24.9	604	1	TYVUH protein-tyrosine k
13	1655.5	24.2	1330	1	GQFPE epidermal growth f
14	1647	24.1	544	2	S35745 protein-tyrosine k
15	1640	24.0	545	2	S00727 kinase-related tra
16	1623	23.7	540	2	B44776 protein-tyrosine k
17	1621	23.7	540	1	TVFVB epidermal growth f
18	1499	21.9	644	2	A63255 protein-tyrosine k
19	1296	18.9	1323	2	B88257 protein let-23 (im
20	1296	18.9	1374	2	S70712 protein-tyrosine k
21	1206	17.6	1369	2	A45558 protein-tyrosine k
22	1159	16.9	1717	1	A42032 epidermal growth f
23	1125	16.4	527	2	A42032 epidermal growth f
24	1000.5	14.6	843	2	A27131 epidermal growth f
25	806.5	11.8	346	2	S13807 protein-tyrosine k
26	754.5	11.0	311	2	T43220 insulin-like growth
27	727	10.6	1363	2	INSUR insulin receptor p
28	708	10.3	1382	1	INSUR insulin receptor p
29	701	10.2	1383	2	A36080 insulin receptor p

30	700.5	10.2	1372	2	A34157 insulin receptor p
31	682.5	10.0	1300	2	A36502 insulin receptor-x
32	681	9.9	1607	2	T43212 insulin-like growth
33	671	9.8	1477	2	T18534 protein-tyrosine k
34	669	9.8	1268	2	B36502 protein receptor-x
35	640	9.3	1367	1	IGHUR1 insulin-like growth
36	629	9.2	1371	2	A33837 insulin-like growth
37	625	9.1	1390	2	T30346 insulin receptor -
38	616	9.0	987	2	A54092 protein-tyrosine k
39	615.5	9.0	2148	1	A56081 insulin receptor -
40	610	8.9	2101	2	S57245 insulin receptor (
41	602.5	8.8	952	2	I50612 protein-tyrosine k
42	597	8.7	987	2	I48652 mouse developmenta
43	594	8.7	984	2	A39753 protein-tyrosine k
44	592	8.6	985	2	I51672 receptor tyrosine
45	591	8.6	987	2	I48953 eph-related recept

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N.Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein et

C.Species: Homo sapiens (man)

C.Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C.Accession: A24571, A25491, A44188, B44188, I59509, I57622

R.Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.

Nature 319, 230-234, 1986

A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth

A.Reference number: A24571, MUID:86118663, PMID:3003577

A.Accession: A24571

A.Molecule type: mRNA

A.Residues: 1-1255 <YAM>

A.Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R.Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid

A.Reference number: A25491, MUID:86016729, PMID:2995967

A.Accession: A25491

A.Molecule type: DNA

A.Residues: 737-1031 <SEM>

A.Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R.Cousens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1132-1139, 1985

A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chron

A.Reference number: A44188, MUID:86070181, PMID:2999974

A.Accession: A44188

A.Molecule type: DNA

A.Residues: 740-910 <COU>

A.Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A.Accession: B44188

A.Molecule type: mRNA

A.Residues: 1-517, 'RALV', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A.Cross-references: GB:M11730; NID:G183986

R.King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A.Reference number: I59509; MUID:85272597; PMID:2992089

A.Accession: I59509

A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 832-909 <REX>

A.Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R.Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcrip

A.Reference number: I57622; MUID:87286898; PMID:3039351

A.Accession: I57622

A.Status: translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AA58637.1; PID:9553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics:
 A:Gene: GDB:ERBB2; NGU: HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 863/3
 A>Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 kinase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:222-653/Domain: extracellular #status predicted <EXT>
 F:770-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>
 F:654-675/Domain: transmembrane #status predicted <TM>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
 F:68/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1232,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6708; DB 1; Length 1255;
 Best local Similarity 98.1%; Pred. No. 1e-267;
 Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELALCRWGLLALLPGAASTVCTGDMKLRLPASPEHLDMLRLHYOGCQVVGQNL 60
 DB 1 MELALCRWGLLALLPGAASTVCTGDMKLRLPASPEHLDMLRLHYOGCQVVGQNL 60
 QY 61 ELTLPPTNASTLSFLDIOEVQGYVLIAHNOVRQVPLQRLRYRGTLFEDYVALAVDNG 120
 DB 61 ELTLPPTNASTLSFLDIOEVQGYVLIAHNOVRQVPLQRLRYRGTLFEDYVALAVDNG 120
 QY 121 DPLNNFNFTVSFWLRVPSKAS----HLEQLRSLLTELKGGVLLQRPOLCYODTLWK 176
 DB 121 DPLNNFNFTVSFWLRVPSKAS----HLEQLRSLLTELKGGVLLQRPOLCYODTLWK 176
 QY 177 DIFKNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDDQSILRTVYAGGACRCKGP 236
 DB 171 DIFKNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDDQSILRTVYAGGACRCKGP 230
 QY 237 LPTDCHEQCAAGCTGPRHSDCLACLAHFNHSGICELHCPALVYNTDTFESMPNDEGRYT 296
 DB 231 LPTDCHEQCAAGCTGPRHSDCLACLAHFNHSGICELHCPALVYNTDTFESMPNDEGRYT 290
 QY 297 FGASCVTACPNYSTDVSGCTLVGCLPLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHL 356
 DB 291 FGASCVTACPNYSTDVSGCTLVGCLPLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHL 350
 QY 357 REVAVTSANIOEFAGCKIFGSLAFLPESPDGASNTAPLOPEQLQVFETLEETITGL 416
 DB 351 REVAVTSANIOEFAGCKIFGSLAFLPESPDGASNTAPLOPEQLQVFETLEETITGL 410
 QY 417 YISAMPDLSPLSVQNLQVTRGRILHNGAYSLSLTQIGISWLGSLRSLRELSGSLALIH 476
 DB 411 YISAMPDLSPLSVQNLQVTRGRILHNGAYSLSLTQIGISWLGSLRSLRELSGSLALIH 470
 QY 477 NTHLCFVITVPMDOFLRNPQALHTANRPEDECGEGELACHQICARHCWGPGTQCVN 536
 DB 471 NTHLCFVITVPMDOFLRNPQALHTANRPEDECGEGELACHQICARHCWGPGTQCVN 530
 QY 537 CSQFLRQGECEVCEGVLOGLPREYVNAHCLPCHBECOPONGSVTCFPEADQCVACAHY 596
 DB 531 CSQFLRQGECEVCEGVLOGLPREYVNAHCLPCHBECOPONGSVTCFPEADQCVACAHY 590
 QY 597 KDPFCVACRCPSSGVKPRDLSYMPIMKFPBEBGACQPCPTNCHSCVDLDDKCCPAEORASP 656
 DB 597 KDPFCVACRCPSSGVKPRDLSYMPIMKFPBEBGACQPCPTNCHSCVDLDDKCCPAEORASP 656

DB 591 KDPFCVACRCPSSGVKPRDLSYMPIMKFPBEBGACQPCPTNCHSCVDLDDKCCPAEORASP 650
 QY 657 LSTISAVNGILLVVGIVGFGILLIKRQOKIRRYTMRLLQETELVEPLTPSGAMPQA 716
 DB 651 LSTISAVNGILLVVGIVGFGILLIKRQOKIRRYTMRLLQETELVEPLTPSGAMPQA 710
 QY 717 QMRILKETELAKVYLGSGAFGVYKGIWIDGENVKIPVAIKVRENTSPKANKELIDE 776
 DB 711 QMRILKETELAKVYLGSGAFGVYKGIWIDGENVKIPVAIKVRENTSPKANKELIDE 770
 QY 777 AYVAVGSPYVSRLLGICLTSTVQLVTQLMPYCGLLDHVRENRGLSGODLLKMCQIA 836
 DB 771 AYVAVGSPYVSRLLGICLTSTVQLVTQLMPYCGLLDHVRENRGLSGODLLKMCQIA 830
 QY 837 KGMVLEDRVLRVHBDLAARNLVSPNNVKTITDGLARLLIDIEVEYADGQVPIKMA 896
 DB 831 KGMVLEDRVLRVHBDLAARNLVSPNNVKTITDGLARLLIDIEVEYADGQVPIKMA 890
 QY 897 LESILRRFTQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICITD 956
 DB 891 LESILRRFTQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICITD 950
 QY 957 VYMIWVKCMIIDSECRPFRELVSFSPMARDPQRFVVIQNEIDLGPASPLDSTFYRSILE 1016
 DB 951 VYMIWVKCMIIDSECRPFRELVSFSPMARDPQRFVVIQNEIDLGPASPLDSTFYRSILE 1010
 QY 1017 DDDMGDLVDAREYLVPOGFFCPCPDPAAGGMVHRRHSSSTRGGGLTGLPSEEEA 1076
 DB 1011 DDDMGDLVDAREYLVPOGFFCPCPDPAAGGMVHRRHSSSTRGGGLTGLPSEEEA 1070
 QY 1077 PRSPILAPSEAGSDVFPDGLMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYAP 1136
 DB 1071 PRSPILAPSEAGSDVFPDGLMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYAP 1130
 QY 1137 LTCSPOPEYVNPQVPRPQPSPREGPLPAARPAGATLERAKTISPGKXGVKDVAFGGA 1196
 DB 1131 LTCSPOPEYVNPQVPRPQPSPREGPLPAARPAGATLERAKTISPGKXGVKDVAFGGA 1190
 QY 1197 VENPEYVLTPOGGAAPQHPAPAFSPADPNLYMOPDPERGAPSTFGCTPAENPEYLG 1256
 DB 1191 VENPEYVLTPOGGAAPQHPAPAFSPADPNLYMOPDPERGAPSTFGCTPAENPEYLG 1250
 QY 1257 LDVPEV 1261
 DB 1251 LDVPEV 1255

RESULT 2
 TYRTNU
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Barrymann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A>Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
 A:Accession: A24562; MUID:86118662; PMID:3945311
 A:Molecule type: mRNA
 A:Residues: 1-1260 <BAR>
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
 R:Maui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen
 Carcinogenesis 12, 1975-1978, 1991
 A>Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals nt
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; MUID:92035293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics:
 A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP, autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domains: signal sequence #status predicted <SIG>
F:120-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domains: transmembrane #status predicted <TMN>
F:723-988/Domains: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:771,191,263,535,576,534/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:758/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.8%; Score 5940; DB 1; Length 1260;
Best Local Similarity 87.0%; Pred. No. 2,9e-235;
Matches 1098; Conservative 50; Mismatches 108; Indels 6; Gaps 3;

1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMLRHLVYGCQVVGNTL 60
4 MELAMCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMLRHLVYGCQVVGNTL 63
61 ELTYLPTNASLFLDIOEVGVYLAHNOVROVLOLRIVRGTOLEFEDNALAVLDNG 120
64 ELTYVPANASLFLDIOEVGVYLAHNOVROVLOLRIVRGTOLEFEDKVALAVLDNR 123
121 DPLNNFNFTVSFWLRVPKVASHLEQLRSLTEILKGVLIQRNPQLCYQDTILMKDIFH 180
124 DPQDVVAATPG--RTPE--GLRELQRLSLTEILKGVLIQRNPQLCYQDVVLMKDVFR 178
181 KNNQALATLITDITNSRACHPCS PMCKGSR CMGESSEDCOSLTRTYCAGGACRCKPLPTD 240
179 KNNQALAPVDITNSRACHPCS PMCKGSR CMGESSEDCOSLTRTYCAGGACRCKPLPTD 238
241 CCHGCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESMNPREGRYTFGAS 300
229 CCHGCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESMNPREGRYTFGAS 298
301 CVTACPPNYVLTSDVSSCTLVCPNLHNOETAEQGTQCEKCSAPCARVCYGLCMENHREVR 360
299 CVTTCPNYVLTSDVSSCTLVCPNLHNOETAEQGTQCEKCSAPCARVCYGLCMENHREVR 358
361 AVTSANIODEFAGCKKIFGSLAFPLPESFPDGPASNTAPLOPELOVPELTETITGLYLSA 420
359 AITSNVOBFDOCKKIFGSLAFPLPESFPDGPASNTAPLOPELOVPELTETITGLYLSA 418
421 WPDLSLDLSVFNLTQVIRGRILHNGAYSLLTQGLGISWLGRLSRLSGSLALIHNTTL 480
419 WPDLSLDLSVFNLTQVIRGRILHNGAYSLLTQGLGISWLGRLSRLSGSLALIHNTTL 478
481 CFVHTVPMDOFLRNHQAHLHTANRPDE--CYGEGLAGHQLCARHNCWGPPTQCVCNSQ 539
479 CFVHTVPMDOFLRNHQAHLHTANRPDE--CYGEGLAGHQLCARHNCWGPPTQCVCNSQ 538
540 FLRGQCEVCECRVLTQGLPREVYNARHCLPCHPECOPOGNSVTCFEPADQCAACHYDP 599
539 FLRGQCEVCECRVLTQGLPREVYNARHCLPCHPECOPOGNSVTCFEPADQCAACHYDP 598
600 PCVYARCSGVKPDLSYMPIMKFPDEBGAQCPINCTHSCVDLDDKCGPAEQRASPTLS 659
599 SSVCARCSGVKPDLSYMPIMKFPDEBGAQCPINCTHSCVDLDDKCGPAEQRASPTLS 658
660 IVSAVAVGLLVVLTGVFGLIKRQOKIRKTKTKRLLQETLVEPLTPSGAMNQAOQR 719
659 IIVAVEGVLFLILLVAVVGLIKRQOKIRKTKTKRLLQETLVEPLTPSGAMNQAOQR 718
720 ILKTELKRVKVLGSGAGFYVKGIMIPDEGVNKIPVAIKVRENTSKANKEILDEAYV 779
719 ILKTELKRVKVLGSGAGFYVKGIMIPDEGVNKIPVAIKVRENTSKANKEILDEAYV 778
780 MAGVGSFYVSRLLGLCLTSTVOLVTLQMPYGLLDHVENRGRGLSODLLNMCQIAJGM 839
779 MAGVGSFYVSRLLGLCLTSTVOLVTLQMPYGLLDHVENRGRGLSODLLNMCQIAJGM 838
840 SYLDVRLVHRDLAARNVLVKSPPNHVKITDFTGLARLLDIDETEHADGKVPITKMALES 899

839 SYLDVRLVHRDLAARNVLVKSPPNHVKITDFTGLARLLDIDETEHADGKVPITKMALES 898
900 IIRRFTHOSDVMSGVTVWELMTGAKPYDGI PAEIRPDLLEKGERLPQPICTIDVYM 959
899 IIRRFTHOSDVMSGVTVWELMTGAKPYDGI PAEIRPDLLEKGERLPQPICTIDVYM 958
960 IMVKCMIDSECRPFRELVSFMSMARDPQRFVYIQNEDLPASPLDSTFYRSLLEDD 1019
959 IMVKCMIDSECRPFRELVSFMSMARDPQRFVYIQNEDLPASPLDSTFYRSLLEDD 1018
1020 MCDLVDAEYVLPQCGFFCPDPAPAGGVNHHRRSSSTRSGGDLTGLPSEEEAPRS 1079
1019 MCDLVDAEYVLPQCGFFCPDPAPAGGVNHHRRSSSTRSGGDLTGLPSEEEAPRS 1078
1080 PLAPSEAGSDVFDGDLGMAKQOSLPTHPSPLOQRSEDPVLPBSETGCVAPLTC 1139
1079 PLAPSEAGSDVFDGDLGMAKQOSLPTHPSPLOQRSEDPVLPBSETGCVAPLTC 1138
1140 SPQPEYVNOQDVRPOPSPREGPLPAAPAGATLSRAKTLSPGKGVNVDVAFGGAIVEN 1199
1139 SPQPEYVNOQDVRPOPSPREGPLPAAPAGATLSRAKTLSPGKGVNVDVAFGGAIVEN 1198
1200 PEYLTPOGGAPOHPHPAPFSPFNLVYWDQDPEPERGAPSTFKCTPTAENPEYGLDV 1259
1199 PEYLTPOGGAPOHPHPAPFSPFNLVYWDQDPEPERGAPSTFKCTPTAENPEYGLDV 1258
1260 PV 1261
1259 PV 1260

RESULT 3
148161
P:185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: 148161
R:Nakamura, T.; Uehlija, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: 148161; MUID:94193007; PMID:7908275
A:Accession: 148161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PIDN:BAA03801.1; PID:9747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domains: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.5%; Score 5921.5; DB 2; Length 1254;
Best Local Similarity 86.7%; Pred. No. 1.6e-235;
Matches 1093; Conservative 58; Mismatches 103; Indels 7; Gaps 3;

1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMLRHLVYGCQVVGNTL 60
1 MELAMCRWGLLALLPFGAASVCTGTDMKRLPASPETHLDMLRHLVYGCQVVGNTL 60
61 ELTYLPTNASLFLDIOEVGVYLAHNOVROVLOLRIVRGTOLEFEDNALAVLDNG 120
61 ELTYLPTNASLFLDIOEVGVYLAHNOVROVLOLRIVRGTOLEFEDKVALAVLDNR 120
121 DPLNNFNFTVSFWLRVPKVASHLEQLRSLTEILKGVLIQRNPQLCYQDTILMKDIFH 180
121 DPLNNFNFTVSFWLRVPKVASHLEQLRSLTEILKGVLIQRNPQLCYQDTILMKDIFH 174
181 KNNQALATLITDITNSRACHPCS PMCKGSR CMGESSEDCOSLTRTYCAGGACRCKPLPTD 240
175 KNNQALAPVDITNSRACHPCS PMCKGSR CMGESSEDCOSLTRTYCAGGACRCKPLPTD 234

```

QY 241 CCEOCACAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPFESMPNPEGRTFGAS 300
DB 235 CCEOCACAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPFESMPNPEGRTFGAS 294
QY 301 CTTACPYNYLSTDVGSCTIWCPLHNOEVTAEADTQCEKSKCPACVCGIAMEHLREVR 360
DB 295 CTTCPYNYLSTFVGSCTIWCPLHNOEVTAEADTQCEKSKCPACVCGIAMEHLREVR 354
QY 361 ATTSANIOEFAGCKKIFGSLAFIPESPDOPANTAPLOEOLOVETLEITGTYLISA 420
DB 355 ATTSANIOEFAGCKKIFGSLAFIPESPDOPANTAPLOEOLOVETLEITGTYLISA 414
QY 421 WPDLSLPLSVFQQLAVIRGRILHNGAYSLTLQGLISWLGRLSRLREISGLALIHNTHL 480
DB 415 WPDLSLPLSVFQQLAVIRGRILHNGAYSLTLQGLISWLGRLSRLREISGLALIHNTHL 474
QY 481 CFVHTVPMQDLFRNPHQALHTANRDEDECVBGLACHOLCARGHCGMPGPTQCVNCSOP 540
DB 475 CFVHTVPMQDLFRNPHQALHTANRDEDECVBGLACHOLCARGHCGMPGPTQCVNCSOP 534
QY 541 LAGOEVEECRVLOGLPREVYNARHCLPCHPECOQNGSVTCRGPADOCVACAHYKDP 600
DB 535 LAGOEVEECRVLOGLPREVYNARHCLPCHPECOQNGSVTCRGPADOCVACAHYKDP 594
QY 601 FCVARCPGKVPDLSTYMPIWKFPEDEGACOPINCTHSCVDLDDKCPAEOGASPLTST 660
DB 595 FCVARCPGKVPDLSTYMPIWKFPEDEGACOPINCTHSCVDLDDKCPAEOGASPLTST 654
QY 661 VSAVVGILLVVLGVVFGILLIKRQOKIRKRYNRLLQETELVEPLTPSGAMPNOQMRI 720
DB 655 IATVVGILLVVLGVVFGILLIKRQOKIRKRYNRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKXKV/GSGAGCTYTKYKIMT/PDGENVKIPVAIKYLRNTSPKANKELLDAYVM 780
DB 715 LKETELRKXKV/GSGAGCTYTKYKIMT/PDGENVKIPVAIKYLRNTSPKANKELLDAYVM 774
QY 781 AGVGSFYVRLTGLICTSTVQTLVQMLPYCCLDHYRENNGRIGSODLNMWQIAKMS 840
DB 775 AGVGSFYVRLTGLICTSTVQTLVQMLPYCCLDHYRENNGRIGSODLNMWQIAKMS 834
QY 841 YLEDVRLVHRDLAARVLYKSPNHVKITDFFGLARLLDIDETEHADGKVPIMWALEST 900
DB 835 YLEDVRLVHRDLAARVLYKSPNHVKITDFFGLARLLDIDETEHADGKVPIMWALEST 894
QY 901 LRRRFTHOSDWSYGVTVWELMFGAKPYDGIPIAREIPDLLEKGERIPQPICTIDVYMI 960
DB 895 LRRRFTHOSDWSYGVTVWELMFGAKPYDGIPIAREIPDLLEKGERIPQPICTIDVYMI 954
QY 961 MWKMMIDSECRPRFELVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLDEDDM 1020
DB 955 MWKMMIDSECRPRFELVSEFSRMAADPQRFVVIQNEDELGPASPLDSTFYRSLDEDDM 1014
QY 1021 GDLVDAEELVLPQGGFFCPDPAPAGQGVHHRSSSTRSGGDLTLGLEPSEEARSP 1080
DB 1015 GDLVDAEELVLPQGGFFCPDPAPAGQGVHHRSSSTRSGGDLTLGLEPSEEARSP 1074
QY 1081 LAPSEBAGSDVPEGDLGMAKAGLOSLPTHDPSPILRYSDDPVVPLPSEEDGVAPLTCG 1140
DB 1075 LAPSEBAGSDVPEGDLGMAKAGLOSLPTHDPSPILRYSDDPVVPLPSEEDGVAPLTCG 1134
QY 1141 POPEVYNQDVRPQPSPREGRLPAARPAAGATLERAKTUSPGNGVGVKDFAFGAVENP 1200
DB 1135 POPEVYNQDVRPQPSPREGRLPAARPAAGATLERAKTUSPGNGVGVKDFAFGAVENP 1194
QY 1201 EYLTPOGGAAPQHPPPAPADNLTYMDQDPEPGAPSTKGTPTAENPEYGLDVP 1260
DB 1195 EYLTPOGGAAPQHPPPAPADNLTYMDQDPEPGAPSTKGTPTAENPEYGLDVP 1253
QY 1261 V 1261
DB 1254 V 1254

```

```

RESULT 4
GOME
epidermal growth factor receptor precursor - human
N:Containing: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A36772; A43615; A23062; A05281; A60143;
R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.
rg, P.H.
A:Nature 309, 418-425, 1984
A>Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <UL>
A:Cross-references: EMBL:X00568; NID:g31113; PIDN:CAA25240.1; PID:g757924
A>Note: the authors translated the codon AAG for residue 540 as Asn
R:Shih, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A>Title: Characterization and sequence of the promoter region of the human epidermal
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M1234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, U.; Whitliffe, N.; Bennett, P.; Kitchington, D.; Ulrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A>Title: The human EGF receptor gene: structure of the 110 kb locus and identificatio
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A>Title: Contributory effects of de Novo transcription and premature transcript termi
A:Reference number: A36772; MUID:91107677; PMID:1988448
A:Accession: A36772
A:Molecule type: DNA
A:Residues: 1-29 <HML>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
R:Lin, C.R.; Chen, W.S.; Krutger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
R.; Xu, Y.; Shih, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;
Nature 309, 806-810, 1984
A>Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCAMRA', 150-187, 'KSVIOAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-
', 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re
R:Lin, C.R.; Chen, W.S.; Krutger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A>Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Stamen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-133, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiser, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

```

R; Russo, N.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A:Reference number: A60143; MUID:85182550; PMID:2985580
 A:Accession: A60143
 A:Molecule type: protein
 A:Residues: 740-744, 'X', 746-747 <RUS>
 R; Mroczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superec
 A:Reference number: A38023; MUID:8491554; PMID:6325948
 A:Contents: annotation; receptor activity
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A:Title: Functional independence of the epidermal growth factor receptor from a domain
 A:Reference number: A33311; MUID:90003233; PMID:2790960
 A:Contents: annotation; internalization signal
 A:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
 C; Genetics
 A:Gene: GDB:EGFR
 A:Cross-references: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-120/Product: EGF receptor #status predicted <KMT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EEL>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F:646-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:148,175,352,413,444,528,603/Binding site: carbohydrate (asn) (covalent) #status predic
 F:745/Active site: Lys #status experimental

Query Match 45.8% Score 3139; DB 1; Length 1210;
 Best Local Similarity 49.3%; Pred. No. 1.5e-121;
 Matches 627; Conservative 179; Mismatches 351; Indels 116; Gaps 22;

```

QY 11 LLLALLPFGAA--STOVCTGDMKLRPASPEYHLMYRHLYOGQVVGNNLELYPTN 68
DB 14 LLAALCPASRALEEKKVCQGTSTNKLQTGTFEDHFLSLQRMNNEVEVLGNLEIYYQRN 73
QY 69 ASLSFLDDIOGVGVLIANHQVROVPLQRLRIYVGTOLFEDNVALLVNDGPIANFN 128
DB 74 YDLSFLKTIQVAGVVLIALNTVERIPLENLIIRGNMYENSVALAVLSYND----- 126
QY 129 FTVSFWLRVPRKVSASHLEQL--RSITELIKGGVLIQRMPOCYOTIIMKDIFFHKNOLA 186
DB 127 -----ANKTGLKELPMRNLOETLHGAVRSPNNPACNVESTIOWRIDVSSDPLSN 175
QY 187 LTLIDNRSRACHPCSPWCKSRGCGESSSEDCQSLTRTVACAGCA--RCKGPLPTDCHEQ 245
DB 176 MSMDFOHNLGSCQKCDPSCPMGSCMGAGEHCQKLTIIICAQCCSGRGRKSGSPSCCINQ 235
QY 246 CAAGCTCPKXBDCLACHFNHSGICELHCPALVYNTDTFESMNPBGRYTFGASCVTAC 305
DB 236 CAAGCTCPRESDCVCRKFRDEATCKDTCPRMLYNPPTYOMDVAVPEKYSFSGATCVKXC 295
QY 306 PYNLTSTDVGSCTLVCPLEHNOEVTAEADGTORCEKSKCAVCVGLGMEHLREVAATVSA 365
DB 296 PRNVVYDHSQVACADSYEM--EDGVKCKCKCEGCRKVCNGIGIGERFDSLSTINAT 354
QY 366 NIOEFACCKIKFGSLAPFESFDDPASPNTAPLOPEQLOVEETLEITGVLYISAMPDL 425
DB 355 NIKFKKCTKSISGDLHLIPVAFRGDSFTHTPRLDQELDIKTYKEITGFLIQAMPENR 414
QY 426 PDLSPFNQVIRIRILHNGAYSLTLQGLISWLGSLRGLGSLALIHNTLCEVHT 485

```

DB 445 TDLHAFENLEIRGRTKQHGQFSLAVSNITSLGRSLKEISDGDVLIISGNRLCYANT 474
 QY 466 VPMDDLFNRPHQALHTNARPEDECYBGLACHQACARHCNGRGTOVCVNGCOFLRGQ 545
 DB 475 INMKCLFGTSGQKTKIISNRGENSCATQVCHALCSPEGCWGPEDVCSCNNVSRGR 534
 QY 546 CVEBCRVLOGLPREVYNARHCLPRHRECOPONGSYTCFPEADQVCACAHYKPPCVAR 605
 DB 535 CYDKCKLLEGEREFEVNSECIOCHRECPROMNITTCRGDNCIOCAHYIDGPHCVKT 584
 QY 606 CPBGVCPDLSTYMPIMKFPDEBACOPCPINCHSCVDLDDKCPABORASPLTISAVV 665
 DB 595 CPAGVWGENVTL-VKXVADAGVCHLCHENCTYGTGCPGLESCPTGPKIP--SIATGMV 651
 QY 666 G---LLVAVVGVNGVGLIKRQOKIRKTKTMRLLQETLVEPLRTSSGAMPNOQRILK 722
 DB 652 GALLLLVALLVGLIG--LFMRNRHIVRKKTLLRLLQERLVBPLTPSGEAPNALLRILK 708
 QY 723 ETELARKVVLGSGAGFVYKGIWIPDGENVKIPVAIKVIRENTSPKANEILDEAYVMAG 782
 DB 709 ETEFKKIKVLSGAGFVYKGLMIEBGEVVKIPVAIKELRENTSPKANEILDEAYVMAS 768
 QY 783 VGSPPVSRLLGLCLSTVQLVLTQMPYGCGLDHDVZENRGLSODLLNMCQIAKMSYL 842
 DB 769 VDNPHVCRLGLCLSTVQLITQMPFGCLLDVREHKONISQYLLNMCVQIAKGMVYL 828
 QY 843 EDVRLVHDDLAARNLVVSPHNVKITTDFGLALLDIDETFYAADGKVPKIMALESLIR 902
 DB 829 EDRLVHDDLAARNLVVTPHNVKITDFGLALLDIDETFYAADGKVPKIMALESLILH 888
 QY 903 RRFTHQSVWVSGVTVWELMTFGAKRYDGIIPAREIPDLLEKGERLPORPICTIDVYMINV 962
 DB 889 RITYHQSDVWVSGVTVWELMTFGSKRYDGIIPAREISLIEKGERLPORPICTIDVYMINV 948
 QY 963 KCMWIDSECRPRFRELVEFSFMRDPORFVVIQ--NEDLGRASPDLSTFYRSLDEDDMG 1021
 DB 949 KCMWIDASRPKRELITIEFSMRDPORYLVIQDERNHLPSPDSNRYRLMDEEDMD 1008
 QY 1022 DLVDAEYLVLPQGGFCDDPRAGAGMNNHRRSSSTSGGDDLTLGLEPSEEARSPYL 1081
 DB 1009 DVVDAEYLVLPQGGF-----SSPSTSRPL 1034
 QY 1082 APSEGAGSDVPDGDGCMGAAGKLOSLPTNDRPRLORYSEDPVPLPSET--DGVAPRLC 1139
 DB 1035 LSSLATSN--NSTYACIDRNGLOSCPIKEDSFLORYSDPGLATEDIDTFL----- 1087
 QY 1140 SPOPEYVQNPVDRPDRPSPREGRLPARAPAGATLERAKTLPKNGGVVDFVAFGAVEN 1199
 DB 1088 -PVEYINQ-SVPRKRPASVQNPVYHNPYLN-----ASRDPHYD--PHSTAVGN 1135
 QY 1200 PEYL-TPOGGAAPDRHPRPAPSPAFDNLTYWDQ-----DP-----DERGAPST 1242
 DB 1136 PEYLVNTVO-----PTCVNSTRPDSPAHMAQKSHQISLDNPYQODFPKKAEPNGI 1186
 QY 1243 FKGPETAENPEYL 1255
 DB 1187 FKGS-TAENAEYL 1198

RESULT 5
 A53183
 C:Species: Mus musculus (house mouse)
 C:Update: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C:Accession: A53183; A43818; S24942; A28941; S45325; I49643
 R:Unetleke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.
 Genes Dev. 8, 399-413, 1994
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
 A:Accession: A53183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <LUE>
 A:Cross-references: GB:U03425

R:Avivi, A.; Lax I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 5, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A:Reference number: A43818, MUID:91232866; PMID:2030916
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eisinger, D.P.; Serrero, G
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, S.K.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941, MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: Protein
A:Residues: 689-694, 'K', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hitbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987, NID:G488830, PIDN:CAA55587.1; PID:G488831
R:Patia, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: 149643, MUID:91126380; PMID:7678348
A:Accession: 149643
A:Status: translated from GB/EMBL/DBU
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
A:Accession: G567201
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphotransferase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3108; DB 2; Length 1210;
Best Local Similarity 49.1%; Pred. No. 276-132;
Matches 629; Conservative 171; Mismatches 360; Indels 120; Gaps 24;

QY 11 LLLALLPFGAA--STVCTGTMDKRLRPASDPETHLDMRLHYOCGVVOCNLTETVPTN 68
Db 14 LTLTALCAAGGALAEKKVCOGTSNRLTQLGTGFEDHFLSLQRYNNCEVLLGNLEITTVGRN 73
QY 69 ASLSFLDIOGVGVVLIHNOVQVPLQRRLRYRGQLPFDNAYALAVLNDGPDLPNNFN 128
Db 74 YDSFLFKTIDGVAGVLLALNTVERIPLENQIIRGNALVENTYALAILSYG----- 126
QY 129 FTVSFWLVRPKVASHTLEQL--RSLTETLKSGVLIQBNPOLCYODTILMKDI-----FKKN 182
Db 127 -----TNRGTRELPMRNLIQELILGAVFNSNNPILCNMDITIOKMDIYGVNFM 175
QY 183 NQALALTLIDNTRSAACHRCSCMCKGSRGWSSESDCSLRTTVACGGA-RCCKGPLTDC 241
Db 176 NISMLP---QSHPSCCPCDPSCPNGSCWGGEENCKLTKTICAQOCSHRCRSPSDC 231

QY	242	CHBOCAGCTBPKNHSDCLACHFNHSGI	CEJHCALYLYNNDTFESMNP	PERGYFGAC	301		
Db	232	CHNOACAGCTBPRESDLVCOXKFODEATC	KOTCPRLMYNPTTOMDVNPEKISFGATC		291		
QY	302	VTACPRVYNLTSDVSSCTLVCPHANOEV	TAEDGTORCEKSSXPCARVCYGLMEHLE	REYRA	361		
Db	292	VKCPRPVYNVVDHDSCVRA	CGPDYEV-EDGIRKCKKCDGPRCKRV	CNGIGIGERKDTLS	350		
QY	362	VTSANIOEFACCKI	FGSLAF	PESFGDDPASPNTAPLOPOLQVETLE	ITYLYISAW	421	
Db	351	INANIKIFKXCTAISGDHLIPAFK	SDSFTRPRPDLPRELEILKTYKEITG	FLLIDAW	410		
QY	422	PDLSLDLSVFNOLVITGRILHNA	AVSLTLOGIGISWGLSLE	BSGLGLIHNNHCLC	481		
Db	411	PDNNTDHAFENLEIRKTYKOHQ	OFSLAVGLNITSGLBSLKEIS	SGDVIISGNRYLC	470		
QY	482	FVHTVPMDOLFRNPHOALLHTAN	PEDECSVEGLACHOLCAR	GHCMPGPPOCVNCSOFL	541		
Db	471	YANTINMKKULFGTNO	KTKIMNNNAEKDCKAVNHVCNPLCS	EGCKMGEPRDVCSCNVS	530		
QY	542	RGQECVEBCRVLOGLPRE	VYNAHRCPCHEPCORONS	VTCFGEADOCVACAHYKDPF	601		
Db	531	RGRECSVECKNILEEBREPE	VENSECTIOCHPECLPRAMNIT	CTGSRGPDNCTIOCAHYIDGPH	590		
QY	602	CVACPSGVKRDLSYMR	IKFPEDEGACQCPERINTHSCVDL	DXKGCBAEQASRLTIV	661		
Db	591	CVKCTCPAGIMENNTL	-VMKYADANNVCHLCHNACTYGC	AGPRGLQCSVWPSGPKI	PSIA	649	
QY	662	SAVIGILLVNLGVNFGI	-LIKRRQOKIRKTYMRLL	LOETELVEITPESGAMPNOAKRI	720		
Db	650	TGIYGLLFIYV	-VALGIGLPMKRRIHVKRTKLRL	DOERLEVEITPESGEPNOAHIRI	708		
QY	721	LKETELRKVLVSGAGFCTYKGI	WIDGENVKI	PVALIKVIRENTSPRANKELIDEAYM	780		
Db	709	LKETEFKKIKYLGSGAFCTYKGI	WIPERGEVKI	PVALIKELREATSPRANKELIDEAYM	768		
QY	781	AGVSPVYSLIGICTST	VOQLVQMPRYGCLLDHVENRGR	LSODLLNMCQIAKMS	840		
Db	769	ASVNPVHCRLGICTST	VOQLVQMPRYGCLLDHVENRKNISQ	UYLLNMCVOIAKMN	828		
QY	841	YLEDVRLVHRLAARNVLVKS	PNHVKTDPFGRLRLDIDET	EYADGGKVRIKMMALESI	900		
Db	829	YLEBRRLVHRLAARNVLVKT	POHVKITDPFGRLALGABEKE	YEAAGKVRIKMMALESI	888		
QY	901	LRRBFTHQSDVMSYGV	TVMELMTFGAKRYDGI	PARREI	PDLLEKGBRLPORPICI	TDVYMI	960
Db	889	LHRITYHQSDVMSYGV	TVMELMTFGSKRYDGI	PASDIS	ISILEKGBRLPORPICI	TDVYMI	948
QY	961	MVKCMIMIDSECRPR	PRELVEFSFSMARDPO	RFVYIQ-NEDUGRASP	LDSTYRSLIEDD	1015	
Db	949	MVKCMIMIDADSRPR	PRELIEFSQMARPO	RYLVYIOGBERNHILSP	PTDSNRYRALIMDED	1008	
QY	1020	MGDVLDAEVLVPOOGF	FCRDPAGAGAMVHNRSSSTR	SGGDLTGLSEPESEARS	1079		
Db	1009	MEDVVDDEYLIPOGFF	-----NSPST-----	SRT	1033		
QY	1080	PLABSEGAGSDVFDGL	CMGA	KGLQSLPTHDP	SPRLQYSEDPV	PLPSET--DGYVAPL	1137
Db	1035	PLLSLSLATSIN	-----NSTVACIN	NGSGRYKEDAF	LQRYSSDPGALV	TEDNIDAFU---1087	
QY	1138	TCSQPREVYNOPDVR	POPSPREBEPPLAAR	PAGATLEBAKTL	SGKQGVKVDVAF	GGAV	1197
Db	1088	---SVPEPVNO-SV	PKRPA	GVQNPVHNOPLR	-----AGRDL	HYON--PHSNVAV	1133
QY	1198	ENPEYL-TPQGA	PARQHPHAPSPAF	DNLYVMO-----DP-----	PERGAPR	1246	
Db	1134	GNEPYLYNTAQ	-----PTCLSS	SGRNSPALMIQK	SHQMSLNDP	YQODFFKXETKEN	1184
QY	1241	STFKGTP	TAENPEYLG	LDVP	1260		
Db	1185	GIFKG-PTAE	NAEYTL	RVAPR	1203		

RESULT 6
TCHLV
Epidermal growth factor receptor precursor - chicken
N.Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C.Species: Gallus gallus (chicken)
C.Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C.Accession: A27720, A00643
R.Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ulrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A.Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A.Reference number: A27720, M01D:88261272; PMID:3260329
A.Molecule type: mRNA
A.Accession: A27720
A.Residues: 1-1223 <LAX>
A.Cross-references: GB:M20386
R.Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A.Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A.Reference number: A00643; M01D:85228222; PMID:2988784
A.Accession: A00643
A.Molecule type: mRNA
A.Residues: 585-1223 <NIL>
A.Cross-references: GB:M10066
C.Genetics:
A.Gene: erbB
C.Superfamily: epidermal growth factor receptor; protein kinase homology
C.Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F.1-30/Domain: signal sequence #status predicted <SIG>
F.31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F.31-654/Domain: extracellular #status predicted <EXT>
F.81-207/Domain: EGF receptor extracellular domain repeat <EEL>
F.197-610/Domain: EGF receptor extracellular domain repeat <EE2>
F.655-677/Domain: transmembrane #status predicted <TM>
F.678-1223/Domain: intracellular #status predicted <INT>
F.719-984/Domain: protein kinase homology <KIN>
F.727-735/Region: protein kinase ATP-binding motif
F.136-202,280,351,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #
F.192,650/Binding site: carboxylate (Ser) (covalent) #status predicted
F.687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F.754/Active site: Lys #status predicted
F.1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.2%; Score 3093.5; DB 1; Length 1223;
Best Local Similarity 48.0%; Pred. No. 1,1e-119;
Matches 626; Conservative 175; Mismatches 351; Indels 151; Gaps 24;

QY 8 RMGLLALLPQGA-----STQVCTGTMKRLRPASPTHLDMRLHYOGCQVVGNDL 61
DB 13 RGAAYLVLLGLGVALCSAVEKKVCCGTNNKLTQLGHEVDHFTSLQRMYNNGCEVLSNLE 72
QY 62 LLYLTNASTSLFLODIOEVGYVLLAHNOVROVPLORLRIIVGTOLEFENYALAVDNGD 121
DB 73 IIVYEHNRDLTKTIOEAGYVLLALANVDVPLENLIQIRGNVLYDMSFALAVALSN-- 130
QY 122 PLANNFNFTVSFWLRVPKVASHLLEOLRSLEILKSGVLIORNPOLCYODTILMKDI 181
DB 131 -----YHNNKIQGLRELPMKRLSELINGVAKISNNPKLCNMDTYLMDIIT 177
QY 182 NNQLALLTLD-TNRSHACPCSPMKSGKSCMGESSEDCSLRTYVAGCA-RCKGPLPT 239
DB 178 SRK-PLTYLDFASNLSSCPKCHPNCNTEHCGWAGGEGNQCOTLTKVICAQCSGRGKIVS 236
QY 240 DCHEGCAAGCTGPKHSDCLCLHFNHSGICELHGFALVTYNTDFEPMNPBGRYTRGA 299
DB 237 DCHNQCAAGCTGPKHSDCLCLCRKFRDATTCKOTCEPLVLYNPTTYQMDVNBGKXSFPA 296
QY 300 SCVTACPNVYSLTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREV 359
DB 297 TCVRCEPNHYVVTVDHSGCVRSQNTDPTVEY-EENGVRKCKCKCGGLGSKVNGIGIGLKI 355
QY 360 RAVTSANIOEFAQCKKIPQSLAFLESPFGDPAASNTAPLOPEOLOVFETLEETGYLYIS 419

DB 356 LSIINATNDSFKNCTKINGDVSLPVALGDAFTKLPDPPKLDVFRVKEISGFLIIO 415
QY 420 AWPDSLPLDSVPONLOVIGRIILHNGAYSLTLOGLGISWGLSLRSLRGLALIHNTH 479
DB 416 AWPDMATDLYAFENLEIRKTRKHQOYSLAVNVLKIOSGLSLKISGDIAMONKN 475
QY 480 LCVHIVPMDQLFRNPHQALHTANRPEDECEGLACHQICARHGCMGPGPOCVNCSQ 539
DB 476 LCADPTMMNRSLFATOSQTKTIQNNKNDCTADRHVCBDLSCDVCGMGFGPHGSCRF 535
QY 540 FLNGDECEVECRYGLPREYVNAHCLPCHPECPORNG--SVTCGPBACDVCACAHY 596
DB 536 FSRQKCEVCQKNILQSEPREFERDSKCLPCHSECLVQNSFAYNTTSGPGPDHCKMAHF 595
QY 597 KQPPFVACPSVCKDLSMPWKPDEGGAQPCPINTHSCVDLDDXGCPAEQASAP 656
DB 596 IDPFCVCAKPAVLGENDTL-VMKYADANAVALCOLCHPNCITRCCKGPGLEGCPR--NGSK 651
QY 657 LRSIVAVY-GILVVVGLVFGILIKRQOKIRKYMRLQETELVEPLTSGAMPNQ 715
DB 652 TSIAGVGVGGLCLVVGIGLGLYLRRL-HYRKATLRLRLQERLEVEPLTSGEAPNQ 710
QY 716 AQMRILKETELRKVYLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPYANKELD 775
DB 711 AHLRIKTEFEKKVYLGSGAFCTVYKGIWIPDGENVKIPVAIKELREATSPYANKELD 770
QY 776 EAYVAVGSPVYSLGICLTSTVQLVQLMAYGCLLDHRENRGLSGQDILNMCQI 835
DB 771 EAYVAVSVNPHCRLLGICLTSTVQLITQLMAYGCLLDVIREHKNIGSQYLLNMCQI 830
QY 836 AKMSYLEDVRLVHRDLAARNVVKSPNHYKIDFGIARLLDDETEYHADGKVPVKM 895
DB 831 AKGMNTEERLVRRLAARNVVKIPONHYKIDFGIARLLDDETEYHADGKVPVKM 890
QY 896 ALESILRRPFTQSDVWSYGVTVWELMTGAKPYDIPAREIPDLLEKGRLEPPRPTI 955
DB 891 ALESILHRYTHQSDVWSYGVTVWELMTGSKYDIPASEISSVLEKGRLEPPRPTI 950
QY 956 DVTMIVKCMIDSEKRPREFVSEFSKARDPQGFVYIQ-NEDLGASPLDSTYRSL 1014
DB 951 DVTMIVKCMIDSEKRPREFVSEFSKARDPQGFVYIQ-NEDLGASPLDSTYRSL 1010
QY 1015 LSEDDMGDLVDAEYLVVPOQGFPCRPDAGAGCMVHNRSSSTRGCGDLTGLERSEE 1074
DB 1011 MEEBDEMDLVDAEYLVVPOQGF-----NSPT----- 1038
QY 1075 EAPRSP-----APSEGAGSDVFDGDLGMAAGLQSLRTHDPSPLQRYSEDPVLPSE 1129
DB 1039 --SRTPLSSLSATSNNSATNCD-----RNGGCHVREDSFVQRSSDPTGNFLEE 1088
QY 1130 T--DGVAPLTCSPQPEYVNOVDVROPQSPREGLPAPARAGATLERAKTSLPGKXGV 1187
DB 1089 SIDDGFL-----PAPEVYNO--LMPKPS-----TAMVONQIY 1119
QY 1188 KDVF-----AFGAVENPEYVLPQGAARPHRPAFAFDLYWDO- 1231
DB 1120 NNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTQNSPLAKTVFESSPYIQS 1171
QY 1232 -----DPE-----RGAPSTFKGPTAEENPEYGLDVP 1260
DB 1172 GNHQINLDNPDVQDFLPNETRPNGLTKPRALENPEYLAANP 1214

RESULT 7
A47253
Epidermal growth factor receptor, HER4 - human
C.Species: Homo sapiens (man)
C.Date: 22-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C.Accession: A47253
R.Plowman, G.D.; Culouesou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A.Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderm

A:Reference number: A47253; MUID:93189574; PMID:8383326
 A:Accession: A47253
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1108 <PLOC>
 A:Cross-references: GB:U07868; NID:9337359; PIDN:AAB59446.1; PID:9337360
 A:Note: sequence extracted from NCBI backbone (NCBI:126842)
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor
 F:716-981/Domain: protein kinase homology <KIN>
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.6%; Score 2986.5; DB 2; Length 1308;
 Best Local Similarity 45.2%; Pred. No. 2.8e-115;
 Matches 612; Conservative 185; Mismatches 374; Indels 183; Gaps 29;

Qy 9 WGLLLALLPPGAA-----STOVCTGTOMKRLPASPETHLMLKHLVGGCGVQVGNLELT 64
 Db WWWVSLVAVAGTVQPSDSQSGAGTENKSLSDLEQVYALRYKYECCVVMGNLEITS 67
 Qy 65 LPTNASTLFLQDIQEOGVYVLIHNNQVQVPLQRLIRVGTOLFEEDNYALAVLDNGDPLN 124
 Db 68 IHNRLDSTFLRSREYVGLVALNORVYLPLENLRIIRGTXYEDRYALAIPLNRYKDG 127
 Qy 125 NFNFTVSEFWLRVPKVSASHLEQ--LRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKN 182
 Db 128 NFG-----LQELGLKNLITEILNGGVYDQNKFLCYADTIHMQDIVRNP 170
 Qy 183 NQALLLIDITNRBRACHPCGPMCKGSRGWSESDQSLRTYACAGC-ARCKGRLPTDC 241
 Db 171 WPSNLTLVSTNGSGGCRCHKSCGTG--RCWGPTEHNCCTLTRYCAECDCRCVGPVSDC 229
 Qy 242 CHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDTFESMPNPEGRTFGASC 301
 Db 230 CHHECGGCGPDTDCFCACMNFNDGACVTCQCPQTVPVPTTFQLEHNNAKTYTAFPC 289
 Qy 302 VTACPVNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 361
 Db 290 VKCKPNFV-VDSSSCVACRPPSSKMEV--EENGIMKCRPCDIPCACDGIQTSIMSAGT 347
 Qy 362 VTNANQEFAGCKKIFGSLAFLEPESFGDDPASNTAPLQPPOLQVFELEITGLYISAM 421
 Db 348 VDSNSIDKFTCKIKINGNLIPLVTGTHGDYNAIEALDPKLVNFRVIREITPELNTQSW 407
 Qy 422 PDSLPDLSVPQNTQVIRGRILHNGAVSLTQGLGSIWLGSRSLRELSGALIHNNHLC 481
 Db 408 PRMTDFSVSNLVTIGRVLVYGLSLILKQOCITSLQSLKEISAGNIYITDNLCL 467
 Qy 482 FVHTVMDQLFRNPHQALLHTANRPEDECVGEGIACHQLCARHGWGPRPTQCVNCSQFL 541
 Db 468 YHTIMVTTLFTINQIRIVIRDRKAKENCTAEGVNCNHLSSDQCGWGPDPQCLSCRFS 527
 Qy 542 RQGECEVEECVLOGLPREYVNAHRCPLRPECQD--QNGSVTCGPREADQVACAHYDDP 600
 Db 528 RGICTIESCNLYGSEFEFENGSIQVCDPQCEMEGGLTTCGPGDNCCTKCSHFQDGP 587
 Qy 601 FCVACRPSGVKPLSYMPIMKFPDEGACOPCPINCTHSCVDDDDKCC-----P 649
 Db 588 NCEKPCPDGLQGANSS--IFKYADPRDECHPCPNCTGCGNGPTSHDCIYPTMGHSTLP 645
 Qy 650 AEOBPASPLTISAVV-GILLVVLGVVFGILLKROQKIRKTYMRLLQETELVEPLTP 708
 Db 646 QHAR-TPL--IAAGVIGGLFVLVIGLTFVAVYARRKIK--KKALRRLFL-ETELVEPLTP 700
 Qy 709 SGAMPNOAMRIKETELRKVKVLGSGAFGVYKGIIPDGENYKIPVALIKVLRENSPK 768
 Db 701 SGAPNOAMRIKETELRKVKVLGSGAFGVYKGIIPDGENYKIPVALIKVLRENSPK 760
 Qy 769 ANKEILDEAYVMAVGSPYVSRLLGLCTSTVQLVTLQMLPYGCLLDHVRNRSGLSODL 828
 Db 761 ANVEFDEALIMSMDBPHLVRLGLVCLSPITQLVTLQMLHGCLELVHVKONISQQL 820
 Qy 829 LNNCMQIAKGMSTLEDEVLVHRDLAARNVLVKSBNHYKITDPGLARLLDITETRYADGG 888

Db 821 LNNCMQIAKGMSTLEDEVLVHRDLAARNVLVKSBNHYKITDPGLARLLDITETRYADGG 880
 Qy 889 KVPKIMMALESLIARRFTQSDVYSGVYVWELMTGAKYDGIIPAREIPDLLEKGRLP 948
 Db 881 KVPKIMMALECHYRRFTQSDVYSGVYVWELMTGAKYDGIIPAREIPDLLEKGRLP 940
 Qy 949 QPPICTIDYVMIMVCKMIDSECRPFRELVSFSSMARDPQRFVYIQNED-LGPASPLD 1007
 Db 941 QPPICTIDYVMIMVCKMIDADSRPFKELAAFSNARAPQRYVLIQSDGRKMLPSBND 1000
 Qy 1008 STFRSLLEDDEDDGDLVDAEYVLPQGFPCPDAPGAGMVHHRHSSSTRSGGDLTL 1067
 Db 1001 SKFPQLLEDEEDLEDDMADEEYLP-QAFNIPPP-----LYTSRARIDNRS-----EI 1048
 Qy 1068 GLEPSEDEAPRS-----PLAP-SRGASDVDPDGLGM 1098
 Db 1049 GHSPPAYTPMSNQFVYRDGCGFAEOGVSVYRAPSTITPEAPVQGAFAELTFDSCCN 1108
 Qy 1099 GAAKGQSLPTHPDPLQRYSEDPVLPSS-----ETDGYVAPLTCSPQPEYVNPQDV 1151
 Db 1109 GTRKRVAPRVQDSSQTSRQADPTVFAPERSRFGELDEEGYMTPRMDRKQOELNVE- 1167
 Qy 1152 RQPPSPREGPLPAPRACATLERAKTLFSGKGVKQVAFGAGAVENPEYLTPOGGAAP 1211
 Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPETHNASNG--- 1194
 Qy 1212 QHPHPPA-----FSPAPNLVYMODPERG 1237
 Db 1195 ---PRADEBYNEPLVNTFANTLGAEVLYKNLISMPKAKKAFONPDYMHSLPPRS 1251
 Qy 1238 A--PSTFKGTP-----AENPEYL 1255
 Db 1252 TLQHPDYLOEYSTKYFYKNGKRIPIVAENPEYL 1285

RESULT 8
 506142
 Nucleotide: tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
 P/Alternate names: epidermal growth factor receptor homolog; kinase-related transform
 C/Species: Xiphophorus maculatus (southern platyfish)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
 C/Accession: 506142; S13809
 R:Mitochondr., J.; Adam, D.; Maltechek, B.; Mauelet, W.; Raulf, F.; Telling, A.; Robe
 Nature 341, 415-421, 1989
 A>Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu
 A:Reference number: 506142; MUID:90015140; PMID:27977166
 A:Accession: 506142
 A:Molecule type: DNA
 A:Residues: 1-1166 <MIT>
 A:Gene: mrk
 A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291
 R:Adam, D.; Mauelet, W.; Scharf, M.
 Oncogene 6, 73-80, 1991
 A>Title: Transcriptional activation of the melanoma inducing xmrk oncogene in xiphoph
 A:Reference number: S13807; MUID:91125882; PMID:1846957
 A:Accession: S13809
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
 A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285
 C:Gene(s):
 A:Gene: mrk
 A:Map position: Y
 A:Insertions: 872/3; 947/1; 979/3; 1025/3; 1056/1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
 F:707-972/Domain: protein kinase homology <KIN>
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 39.1%; Score 2679; DB 1; Length 1166;
 Best Local Similarity 45.0%; Pred. No. 9.7e-103;

Matches: 575; Conservative: 163; Mismatches: 387; Indels: 152; Gaps: 29;	
Qy	4 AALCRWGLLALIPGAAST---QVCTGDMKRLRPASPTHLDMLRHL.YOGCQVYQGN 59
Db	8 AALLQ--LLLVLSIRSCSTDPDRKVCQGTSMQTM---LDNHYLKKMKXVSSCQNVLEN 62
Qy	60 LELTLYPTNASTLFLDIOIEVQGYVLI.IAHNOVRQVLRQLRIYRGQLPEDNYALAVLDN 119
Db	63 LEITTYOENODLSFLDSIQEVGGYVLI.IAHNEVSTIPLVNLRLRGQNLVEGNTLLYMSN 122
Qy	120 --GDP.LNNFNFTVSEFMLRVPKVASHLBOLR--SLTEILKGGVLIQRNPOLCYODTILM 175
Db	123 YQKNSS-----PDVYQVGLQQLQSLNLTSLSSGVKXSNHPILCNVETIM 169
Qy	176 KDI.FHNKNOLAL.TLIDTNRSRACHPCSPMKCKSGRCWSESDCQSLTRTVACAGC--ARCK 234
Db	170 WDIVKTSNPTMMLLP.HAFEROCQCKDHCCVNGSCWAPRGHCQKFTKLLCAEQCRRCR 229
Qy	235 GLPLPDCCHCECAAGCTGPKHSDCLALPHNSIGTIELNCPALVTYNTDTFESMPREGR 294
Db	230 GKPLIDCCNEHCAGCTGPRATDCLACRDFNDGTCOTCPPKXIYDIYSHQVNDPNIK 289
Qy	295 YTFGASCVTACPYNYLSTDVSGCTLVCP.LHNOEVTAEADGTORCEKCKSPCARCYGLGME 354
Db	290 YTFGAACVCEPSNYVTE--GACVRCSCSAGMLEVD--ENGRKCKPCDGVCPKVCDSIGIG 347
Qy	355 HLREVRATVSANI.QEFAQCKKI.FGSLAF.PESFDGPASNTAPLOEOLQVETLEITG 414
Db	348 SLSTNTIAVNSTIRSFNSCTKINGDII.LRNSFEGDPHYKIGTMDPEHLMNLTVEITG 407
Qy	415 YLYISAMPDLSLDLSFONLOVIRGILHNGAYS--LTLOGLSIWSLGLSLBELSGLAL 473
Db	408 YLVIMWPMNMTSLSPQNLLEIRGHTFSRGSFVYVQVRHLQWGLSLKLEKVSAGNYI 467
Qy	474 IHNHNLGCVHTVPMDQLFRNPQALHTANRPEDECEGLACHOLCARGCWGRGPTQ 533
Db	468 LKNTIQLRVANTINNRRLFRSEDSQIEYAR-----ENOTCNNECEDGCM--PGPTM 519
Qy	534 CYNCGQLFLRGCEVCEBRLVQL.PRYVNAHCLPCHPECPQNSVTCFGPADCVAC 593
Db	520 CVSCHLHVRGRCVASCNLLQCEPREAOVDGRVCVQCHQCLVQTSILCYGGRANCSKS 579
Qy	594 AHYKDPFCVACPSGVKNDLSYMPKWPDEEGACQRPINCHTSVULDDKCPAEGR 653
Db	580 AHFOGQPCIPRCFPHGILGDDPTL--IMKADYKMGQOQPHOHCCTQCSGPGLSGRGD--I 637
Qy	654 ASPLTSISAVVGI.LLVVLGV.FGILLKRRQOKIRKTYMRRLQETELVEPLTPTSGAMP 713
Db	638 VSHSLAVGLVSGLLITVIVALLIVLRRRIK--RKRTIKCLLOEKELVEPLTPTSGAP 696
Qy	714 NOQMRLIKETELRKVKVLGSGAFGTVYKGIWI.PQGENYKIFVALKVLRENTSPYANKEI 773
Db	697 NOAFRLIKETEFKKDRLVLSGAFGTVYKGLWNPGENIRIPVALKVLRENTSPYANKGV 756
Qy	774 LDEAVVMAVGSPVYSLIGICTSTVQVLTQVLMRYGCLLDHVRNENRGLSGODLLNMCW 833
Db	757 LDEAVVMAVSDHPVHCRLIGICTLSAVQVLTQVLMRYGCLLDVROHORIICQWMLNMCV 816
Qy	834 QIAKMSYLEDVRLVHRDLAARNV.LKSPNHVKTIDFGLARLDDIDETRYHADGKVPIK 893
Db	817 QIAKMNVLLEERHLVHRDLAARNV.LKSPNHVKTIDFGLSKLLTDDEKEYQADGKAVPIK 876
Qy	894 WMALESILRRRFTHOSDVWSYGVTVWEIMTFGAKPYDGI.PAREIPDLLEKGRRLPOPPIK 953
Db	877 WMALESILQWYTHOSDVWSYGVTVWEIMTFGSKPYDGI.PAKEIASVLENGERLQOPPIC 936
Qy	954 TIDVYMIWKCMYIDSECRPRELVSFESMAAPRQCFVYIQNDLGPASLSDTFPRS 1013
Db	937 TIEVYMIILKCMYIDPSRPRELVSFESMAAPRQCFVYIQG---NLPSLSDRLPSR 993
Qy	1014 LLEDDMDLVDAEELVLPQGFPCPDAPAGAGWVHHRSSSTRSGGDLTLGLPSE 1073
Db	994 LLSDD--DVVDADDEYLLPYKRI-----NRQS----- 1019

Qy	1074 EEARSPPLAPSEAGSDVFDGLGMAKGLQSL.PTHDPSLPORYSEDPV--PLPSETDG 1132
Db	1020 -----EPCIPPTGH-----PRENSITLRINISDPTQMLEKOLD 1054
Qy	1133 YVAPLTCSQPEVYNQPDVRRPQ-----PSPRE-----GLLP--AARPGATLERAKTL 1179
Db	1055 H-----EYVNPQGETSSRLSDIYVNPEDLTDGWPVLSLSQEAETNFSRPEYL 1104
Qy	1180 SPQKGVVVDVAFAGVAVNPEYLTPOGGAAAPPHPRPAPSPAFDNLVYWDQPPRENGAP 1239
Db	1105 NTKQNSL---PLVSSGMDDPDY---QAG-----YQAF-----LPQYGL 1139
Qy	1240 PSTFKGTPTAENPEYLG 1256
Db	1140 TQNGMFLPAENLEYLG 1156

RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epiderma
A:Reference number: A36223; PMID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Flowman, G.D.; Whithney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, J.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor
A:Reference number: 159164; PMID:90311312; PMID:2164210
A:Accession: 159164
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:9183990; PIDN:AAA35979.1; PID:9306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
A:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.3%; Score 2418.5; DB 2; Length 1342;
Beet Local Similarity 40.4%; Pred. No. 5.2e-92;
Matches 533; Conservative 192; Mismatches 454; Indels 141; Gaps 34;

Qy	10 GLLALLPFGAA--STOVCTGDMKRLRPASPTHLDMLRHL.YOGCQVYQGNLELYLPT 67
Db	11 GILFSLANGSEVNGQAVCPGTLNGLSTVGAENQYQTLTKYKERCEVVMGULEYLTGH 70
Qy	68 NASLSFLDIOIEVQGYVLI.IAHNOVRQVLRQLRIYRGQLPEDNYALAVLDNGDPLANNF 127
Db	71 NADLSFLQWIREVTSYVLI.VANNEFSTLP.PNLRVVRGTVQVQKFAIFV-----MLANN 124
Qy	128 NPTVSMFLRVPRVNSH--LEQLR--SLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNQ 184
Db	125 -----TNSSHLRQLRLTQLTEILISGGVYIKNDLCHMDTIDMDIVDRD-- 171
Qy	185 LALTLIDTNRSRACHPCSPMKCKSGRCWSESDCQSLTRTVACAGC--ARCKGLPPTDCH 243
Db	172 --AEIVVDNDRSCPCHEVCKG--RCWPGSEDCQTLTKITCAIPCQNGCHCFEPNQCCH 228
Qy	244 EQCAAGCTGPKHSDCLALPHNSIGTIELNCPALVTYNTDTFESMPREGRYTFGASCVT 303

```
Db 229 DECAGCGSGPQDIDCFACRHNFDNSGACVRCPOPLVYNNKLTQLEPNPHTKYOYGGCVVA 288
QY 304 ACPYNYLSTDVSGCTLLVCPRLHNOEYTAEBGTORCEKCSKPCARVCVGLGMEHLREBRAVT 363
Db 289 SCPNHFV-VDTQSCVRACPPDMEVD-KNGLKMKCECGGLCPACGCTGSG--SRFQYD 344
QY 344 SANIOEAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVEFTEETITGYLYISAMPD 423
Db 345 SSNIGFVNCTKILGNLDFLITGLNDPNNHKIPALDEPLKNTFRTYRELTGNTLQSWRP 404
QY 424 SLPLDSVFONLOVIRGRILHNGAYS-LTLOGIGISWGLRSJRELGSGLAIHHNTHLCE 482
Db 405 HHMNFVSFNSLTITIGRSLYNKGFSLLIMKNLVNLSLGFRRSLKELISAGRIYISANQLCY 464
QY 483 VITVPMDOLEFRNPHOALLHTA-NRPEDECVBEGGLACHOLCARGHGCMGPPTOCVNSOGL 541
Db 465 HSLMTWKVLRGPTERRLDIKNNRPRRDCVABGKVDPLCSSGCGGPGGGGLSCRNYS 524
QY 542 RGOECVEECRVLOGLPREYVNAARHCLPCHPECOPONGSVTCFQPEADOCVACAHYDPPF 601
Db 525 RGVGVCTHGNFLNGEPRFAHEACFSCHPECOQPMGSTATCNGSGSDTCAQCAHPFDGPH 584
QY 602 CVARCPGCVKPLSTYPIKFPDEBACQCPPTNCTHSCVVDLDDKCCPAEORA---SPL 657
Db 585 CVSSCPHVLG--AKGPITYKPDVONECRPCHENCTQCGKGPEDQLGQTLVLIGKTHL 642
QY 658 TSIVSAVVGILVVVGVVFGILIKRROOKIR-KYTMRRLOTELVEPLTPESGAMPNOA 716
Db 643 TVALVTIAG--LVVIFPMGLGTFLYVRGRIQNKRMRYLEGGESIEPLDPS-EKANKV 699
QY 717 QMRILKETELRKVKVLSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPRANKELDE 776
Db 700 LARIFRETETRLKVLGSGVFGTVHKGVMIPEGESIKIPVCIKVIDKSGRQSFQAVTDH 759
QY 777 AVYAVAGSVYVSRLLIGLITSTVQLVTOQLMPGCLLDHYRENRGLSGODLLNMCQIA 836
Db 760 MLAIQLDHAHIVRLGLCPGSSLDQVLYLPLGSLDDHYRQHRGLGQLLLMVOYA 819
QY 837 KCMASYLEDVRLVYRDIAARNVLKSPNHYKITDFGLARLLDIDETEYHADGGCVPIKMA 896
Db 820 KGMYLEHEGHWNRILAAARNVLKSPQOVADFGADLLPDDKLLYSEAKTPIKMA 879
QY 897 LESILRRRTFTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTID 956
Db 880 LESIHGKYTHOSDVMSYGVYVWELMTFGAEPYAGLRLABVPLLEKGERLAQPICTID 939
QY 957 VYMIWVKCMIMISECCPRPRELVSESRAPDRQRFVIVQNEELGRA---SPLDSTFYRS 1013
Db 940 VYMWVKCMIMIDENIRTEKELANETTRMARDPRLVLIKRES-GRGIAPGPEPHGLTNK 998
QY 1014 LLEDDMDGLVDAEYLVFQCGFFCDDPAFGAGCMVHHRRSSSTRSGGDLITGLEP-S 1072
Db 999 KLEEVLEBELDLDLEAED-----NIATTTLGSLSLPVTGIAN 1039
QY 1073 EEEAPRSPLAPSEGASDVFDGDLGKAAGLQSLPLTH-PSPLQRYSEDPVPLP--- 1127
Db 1040 RPRGQSLLSPSSGY-MPMNQNLIGESCQESAVSSESRCPREVSLH-----PMRGL 1092
QY 1128 --SEPDGYVA-----PLTCSQPE---YVNPDRVQPPSPREBP--- 1162
Db 1093 ASESEGHVTSGEAEIOEVSMCRSRRSRPRPRDSAYHSQHSLLTPVPPLSPGLE 1152
QY 1163 -----LPAARPAATLERAKTSLP-GKNGV-----KDFAFAGAVENPEYLTPOG 1207
Db 1163 EEDVNGVYVMDTHLNGTSPSSREGTLLSVGLSVLTGEEDEED-----EEYVYNNRR 1204
QY 1208 GAAPORPPAPSPAPADNLYYMD-----ODPBRGAPSTFYKTPAENPEYL 1255
Db 1205 RHSP-RHPRPSSLLELGYEYMDVGSGLSASLGSTQSCPLHPVPIMPITAGTTPDEDEYEM 1263
```

RESULT 10
JC4387

```
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: J04387
Gene: Heliyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protei
A:Reference number: J04387; MUID:96096555; PMID:8522190
A:Accession: J04387
A:Molecule type: mRNA
A:Residues: 1-1339 <HELV>
A:Cross-references: GB:U29339; NID:G915389; PID:G915390
A:Experimental source: liver
A>Note: The authors translated the codon AAT for residue 369 as Thr and GAT for resid
C:Comment: This protein is a functional heregulin receptor that transduces signals to
C:Genetic:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1339/Product: epidermal growth factor homolog #status predicted <Mat>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <Kin>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 34.1%; Score 2334.5; DB 2; Length 1339;
Best Local Similarity 40.5%; Pred. No. 1,4e-88;
Matches 523; Conservative 173; Mismatches 429; Indels 167; Gaps 36;

QY 3 LAALCRWGLLALLPFGAA---STOYCTDMDKRLRPAPEHMLDLRLHYOGCOVVOGN 59
Db 7 LOVLC-----FLSLARGSMGNSQAVCPBTGLNGISTGADNOYOTLLYLYKECEVVMGN 62
QY 60 LELVTLPTNASLDFIODEVOGYVLIANOVROYVLRKIVRGTLGFEDNYALAVLDN 119
Db 63 LEIVLTGNHADLSFIQWIREYAVLVANNEFSVLPRLRVRRGVGYVQDGKFAIFV--- 119
QY 120 GRLPLNNFNFTYSFMLRVKVSASH-LEQDR--SLTEILLKGVLIQNRNOLCYODITLWK 176
Db 120 ---MLNLYN-----TNSSHALRLKLTQLELLISGGVYIEKNKLCMDITIDWK 164
QY 177 DIFHKNNQALTLIDTNRGRACHPCSPMKGSGSESESDCSLRTYVCAAGC-ARRKG 235
Db 165 DIVRYR---GAEIVKNNANCPGHEVCKG-RMGPRGDDCOILKTKICAPCNGRCRG 220
QY 236 PLPTDCHEQCAAGCTGPRHSDCLACLAHFNHSGICELCAPALVTYNTDIFESMPNEGRY 295
Db 221 PNPNOCHDECAAGCGGPQDIDCFACRHNFDNSGACVRCPRPEPLVYNNKLTQLEPNPHTKY 280
QY 296 TFGASGVTACPNNYISTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLMEH 355
Db 281 QYGVGVASCPNHFV-VDTQFCVRACPPDMEVD-KNGLKMKCECGGLCPKACEGSGG- 337
QY 356 LREVRVNTSANIOEAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVEFTEETITGY 415
Db 338 -SRVQVDSNSNIDGFNCTKIKLGNLDFLITGLAVDPMHKIPALDEPLKNTFRTYRELTG 396
QY 416 LYISAMPDSLPLSVFONLOVIRGRILHNGAYS-LTLOGIGISWGLRSJRELGSGLAI 474
Db 397 LNIQSWPPIHMHFVSNSLTITIGRSLYNRGFSLLIMKNLVNLSLGFRRSLKELISAGRYI 456
QY 475 HNHTHLCPNHYTPMLOLFRNPHOALLHTA-NRPEDECVBEGGLACHOLCARGHGCMGPPO 533
Db 457 SANQQLCYHNSLNMWRLRLGPEERLDITRYDRPLGCLABGKVCDDLCSGGCGWGAPOQ 516
QY 534 CVNSGQFLRAGCEVEECRVLOGLPREYVNAARHCLPCHPECOPONGSVTCFQPEADOCVAC 593
Db 517 CLSCNNYSRSGCVVTHGNFLNGEPRFAHEACFSCHPECOQPMGSTATCNGSGSDTCAQCA 576
QY 594 AHYKORPFCVVARCPGKPLSTYPIKFPDEBACQCPPTNCTHSC--VDLDDKCCPAE 651
```

Db 577 AHFRGPGICVNSCPHGLG--AKGPIYKTPDQNECRPHENCTQCCNGPELDLCLQAE 634
Qy 652 QASPLTSIVSAVVGILLVVLGVFGILLIKRQOKIR-KYTMRLLOETELVEPTPSG 710
635 VLMSKPHLVIAVTG--LAVILMIIGSFLYWRGRIRQKRAMRRLLENGESIEPLDPS- 691
Qy 711 AMPNOAKRILKETELRKVKVLGSGAFGTIVYKGIIPGSENVKIPALIVLENTSPKAN 770
692 EKANKVLRIKFKETLRKLVKLGSGVFGTVHKGIIPGSEIKIPVICIVIDKSGROSF 751
Qy 771 KEILDEAVYMGVSPVYSRLIGICLTSTVOLVTQMLPGCLLDHYRENRGLSGSDLLN 830
752 QAVTHMLAVGSLDPAHIVRLGLCPGSSLDVLYTLPGLSLDHYKQRETLGPPOLLN 811
Qy 831 WCMQIAKGMSTYLEDVRLVHRDLAARNVLKVSBNHYKITDFGLARLLDIDETEYHAGSKY 890
812 MGVOIAKGMVYLEHSVMYHRDLARNVLMKSPSQVADFGVADLLPDDKQLHSEAKT 871
Qy 891 PIKMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPP 950
872 PIKMALESIHFGKYTHOSDWSYGVTVWELMTFGAPYAGLRLAEIPDLLEKGERLAP 931
Qy 951 PICTIDVYMIWCMIMIDSECRPFRELVSFSRMAKDPQRFVVIQNEDLGASPLDSTF 1010
932 QICTIDVYMIWCMIMIDENIRPTRELANETKMAKDPRLVLIKAS-GEQTP--PAA 988
Qy 1011 YRSLEDDMDGLVDAEYLVPOQGFCDPPAPGAGVHHRRSSSTRSGGGLTLGLE 1070
989 EPSVLTKEL-----QEAELEPEL-----DLDLDE 1014
Qy 1071 PSEF-----EAPNSPLAPSE-----AGSDVFGDDLGMGAAK 1102
1015 AEEELATSLGSLALPTGTLTRPGSOSLSPSSGYMPMNSLGEACLDAAVLGREQ 1074
Qy 1103 GLQSLPTHPDPSLQRYSEDPTVPLPSETDGYV---APL-----TC-----SPOPE-- 1144
1075 FSRPISLH-PIPRGR-----PASESSGHVITGSEAEIOEVSVCRSRSRSPRPGD 1126
Qy 1145 --YVNPDPVRPOPSPREGP-----LPAARPAAGATLERAKTLP--GKNGV-- 1187
1127 SAYHSQSRSLTLTPVPLSPGLEEDNGVMPDTHLRGASSREGLTSSVGLSYLQTE 1186
Qy 1188 ---KDVAFAGAVENPEYLTPOGGAAPQHP 1216
Db 1187 EEDED-----EEYEYMNKRKRGSP--PRP 1209

RESULT 11
TVFVY
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Matoney, P.A.; Goodwin, R.G.; Rottman, F.M.; Citterenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NT>
A:Cross-references: GB:M10066; GB:M13801; NID:g211749; PIDN:AAA48763.1; PID:g211750
C:Note: in Genbank entry CHKRBFR, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ENK>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted
Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 1,4e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
Qy 584 GPEADQVCACAHYKDPFVCARCPGCVKDLSTYMPIKRPDEEGACOPCEINCHSCVDL 643
60 GP--DHCMCAFIIDGPHCVKACPGVGLGENDTL--VMKYADANAACQLCHPNTCRCKGP 116
Qy 644 DDKGPAERASPLTSIVSAV--GILLVVLGVFGILLIKRQOKIRKXTMRLLDTEL 702
117 GLEGGP--NGSKTBSIAGVVGGLCLVAVGIGLYLRR--HIVKRTLRLLDEREL 172
Qy 703 VEPLTPSGAMPNOAKRILKETELRKVKVLGSGAFGTIVYKGIIPGSENVKIPALIVLENTSPKAN 762
173 VEPLTPSGAMPNOAKRILKETELRKVKVLGSGAFGTIVYKGIIPGSENVKIPALIVLENTSPKAN 762
Qy 763 ENTPSPKANKIIDEAVYMGVSPVYSRLIGICLTSTVOLVTQMLPGCLLDHYRENRGR 822
233 EATSPKANKIIDEAVYMGVSPVYSRLIGICLTSTVOLVTQMLPGCLLDHYRENRGR 292
Qy 823 LGSOLLWCMQIAKGMSTYLEDVRLVHRDLAARNVLKVSBNHYKITDFGLARLLDIDETE 882
293 IGSOYLLWCMQIAKGMSTYLEDVRLVHRDLAARNVLKVSBNHYKITDFGLARLLDIDETE 352
Qy 883 YHADGKVPKIMMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 942
353 YHADGKVPKIMMALESILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 412
Qy 943 KGERLPOPICTIDVYMIWCMIMIDSECRPFRELVSFSRMAKDPQRFVVIQNEDLGASPLDSTF 1001
413 KGERLPOPICTIDVYMIWCMIMIDADSRPFRELVSFSRMAKDPQRFVVIQNEDLGASPLDSTF 472
Qy 1002 PASPLDSTYRSLEDDDDGLVDAEYLVPOQGFCDPPAPGAGVHHRRSSSTRSG 1061
473 LPSPTDSKRYRTIMEEDMEDIVDAEYLVPOQGF-----NSPST-- 513
Qy 1062 GGDLTGLLEPSEEAAPRPL-----APSEGASDVVDGLGMGAAGLQSLPTHPDPSPLQ 1116
514 -----SRITLLSSLSATSNSTNCID-----RNGGHPREDSFVQ 550
Qy 1117 RYSEDPTVPLPSET--DGVAALTCSPPOPEYVNPDPVRPOPSPREGPPLPAARPAAGATLE 1174
551 RYSDPTSGNGLFRESIDGFL-----PAPEYVNO--LMPKXS----- 585
Qy 1175 RAKTLPKAGVYKVF-----AFGAVENPEYLTPOGGAAPQHP 1219
586 ---TAMVONQIYNNISLTAISKLPWDSRYQNSHTAVDNPEYL-----NTNOSPILA 633
Qy 1220 SPAPNLLYWDQ-----DPPE-----RGAPSTKGTPTAENPEYLGIDVP 1260
634 KTVFESSPYIWSGNHQLNDNPYQDPLPNETYKNGILKVPAAENPEYLRVAAP 689

RESULT 12
TVYU
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
C:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fa
A:Reference number: A00644; MUID:84026539; PMID:6133229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:9209676; PIDN:AAA42400.1; PID:9209678
R:Debnire, B.; Henry, C.; Benaissa, M.; Bisette, G.; Claverie, J.M.; Saulie, S.; Martin
Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type
A:Reference number: A38022; MUID:84223957; PMID:6328658

A:Accession: A38022
 A:Molecule type: DNA
 A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
 A:Cross-references: GB:K02006
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F:130-395/Domain: protein kinase homology <KIN>
 F:138-146/Region: protein kinase ATP-binding motif
 F:165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;
 Best Local Similarity 52.2%; Pred. No. 4, 8e-63;

Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```

Qy 593 CAHYKDPFCVAPCGSVKPDLSYMPWKPEDEGACQPCPINCYSQVLDKDCPRAEQ 652
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VRKYADANAVCOLCHPNCCTGCKGPGCEGP--- 58
Qy 653 RASPLTSISAVV-GILAVVVLGVVGLIKRQOKIKRYTMRLLQETELVEPLTPSGA 711
Db 59 NSGKTPSIAGVVGGLCLVWVGLGILVLRK-HIVRKKTLLRLLQERLEVEPLTPSGE 117
Qy 712 MENOQMRILKETELRKVKVLSGAGFVYKGIWIPDGENVKIPVAIKVRENTSPKANK 771
Db 118 APRQALRLIKETEFKVKVLSGAGFTIYKGLMIPGEKVKIPVAIKELRENTSPKANK 177
Qy 772 EILDEAYVAVGSPVYSRLIGLITSTVQLTQLMPEYGLLDHVENRGLSGODLLNM 831
Db 178 EILDEAYVAVSVNPHVCRLGLITSTVQLTQLMPEYGLLDHVENRGLSGODLLNM 237
Qy 832 CMQIAGMSYLEVRLVHRLAARVLYKSPNHVKITDPELADLLDDEFEYADGKVP 891
Db 238 CVQIAGMNYLERRLVHRLAARVLYKTPQVHKITDPELADLLDDEFEYADGKVP 297
Qy 892 IKMALESILRRFTQSDVMSYGVYWEIMTFGAKPYDGIIPAREIPDLLEKGERLPQRP 951
Db 298 IKMALESILHRITHTQSDVMSYGVYWEIMTFGSKPYDGIIPAREIPDLLEKGERLPQRP 357
Qy 952 ICTIDYVMIMVKCMWIDSECRPRFELVSEFSRMDRPORFVYIQ-NEDUGPASPDLSTF 1010
Db 358 ICTIDYVMIMVKCMWIDSECRPRFELVSEFSRMDRPORFVYIQ-NEDUGPASPDLSTF 417
Qy 1011 YRSLEDDDDMGDLVDAEYVLVPOGFCPPDPAPGAGMWHRRSSSTRSGGDLTLGLE 1070
Db 418 YRSLMEEDMEDIVDAEYVLVPHOGFF-----NSPST----- 449
Qy 1071 PSEEGAPRSPL-----APSEAGSDVDFDGLGMAKGLQSLPTHDPSPLOQRYSEDPVTP 1125
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGGHPVREDSFVQRYSSDPTGN 495
Qy 1126 LPSET--DGVALTQSPQPEYVNOQRPVRQPPSPRGRPLPAARFAGATLERAKTSLPGK 1183
Db 496 PLEESIDDGFL-----PAPEYVNO--LMPKRPSTAN----- 524
Qy 1184 NGVYKOVFAF-----GAVENPEYLTTPGGGAAPQHPAPAFSPAFD 1224
Db 525 --VQNDIYNFISLTATSKLPMDSRYQNSHSTAVDNPEYL-----NTQSPFLAKTVFE 574
Qy 1225 NLVYWDQDPPERGAPSTFGKTPAENPEY 1254
Db 575 SSPLYWIOSGNHQ-----INLDNPQY 594

```

RESULT 13

epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C:Species: Drosophila melanogaster
 C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
 C:Accession: A00640; A38021
 R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.

Cell 40, 599-607, 1985
 A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone bindi
 A:Reference number: A00640; MID:85124611; PMID:2982499
 A:Molecule type: DNA
 A:Residues: 1-1330 <Liv>
 A:Cross-references: EMBL:K03054
 R:Madsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
 Nature 314, 178-180, 1985
 A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor
 A:Reference number: A38021; MID:85137938; PMID:2983232
 A:Accession: A38021
 A:Molecule type: DNA
 A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLYK', <MAD>
 A:Cross-references: EMBL:X02293; MID:g7922; PIN:CAA26157.1; PID:g929565
 C:Comment: This sequence is tentative because the introns have not been identified.
 C:Genetics:
 A:Gene: FlyBase:Bgfr
 A:Cross-references: FlyBase:Fgdn0003731
 A:Map position: 2 57F
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
 F:1-732/Domain: extracellular #status predicted <EXT>
 F:733-764/Domain: transmembrane #status predicted <TM>
 F:765-1330/Domain: intracellular #status predicted <INT>
 F:808-1072/Domain: protein kinase homology <KIN>
 F:816-824/Region: protein kinase ATP-binding motif
 F:122-300,324,363,518,688,695,700/Binding site: carbohydrate (Aen) (covalent) #status
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predi
 F:843/Active site: Lys #status predicted
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pred

Query Match 24.2%; Score 1655.5; DB 1; Length 1330;
 Best Local Similarity 29.9%; Pred. No. 9, 2e-61;

Matches 414; Conservative 180; Mismatches 414; Indels 377; Gaps 40;

```

Qy 80 VQGYVLIANQVQVPLQRIYRGTOLE-----EDVYALAVDNGDPLNFFNFTGFW 134
Db 38 ITNYIYIGDLICTLSYRQIIRGRLPSLSBEERYAL-----FVYISMYT 86
Qy 135 LRVKVASASHLEQLRSITELIKGCVLIQRNPOLCYODITLMKDI FHKNNOLATLTLIDTNR 194
Db 87 LEIP-----DLRVLNQGQVGFHNNYNNLCHMRTIQMSEIVSNGTDAVYVDTAP 135
Qy 195 SRACHCSPKCKSRCKMGESSEDDQSILTRVACAGCA--RCKGLPDDCHGCAACTG 252
Db 136 ERECPKCHESCTHG--CMGEGPKNCQKFSKLTCSPPQCGAGRCYGRPRECHLFCAGGCTG 194
Qy 253 PKASDCLACLHFNHSGICELHCPALVYNTDFTESMNPREGRYTFGASCVTACPYNYLST 312
Db 195 PTKDCIACGNFDEAVSKCECPMKRYNTTLYVLEINPBGKAYGATCYKECP--GHLLR 253
Qy 313 DVGSCITLVCPLHNOEYTAEDGTORCEKSPCARVCYGLAMEHLREYRAVTSANIOEPAG 372
Db 254 DNGACVRSCTPDQKMDKGE-----CVPCNGPCPCTCGVTVLH-----AGNIDSFN 300
Qy 373 CKKIFGSLAFPSFPG--DPAANTA-----PLQEQLOVFELEITGYLYISAMPDS 424
Db 301 CTVIDENIRILDTQTFGFDVVVANYTWGPRYIPLDPERRNVFSTVKEITTYLNEGTHPQ 360
Qy 425 LPLDSFQNLQVIRGRILHNHAY--SLTLQIGISWILRLREISGLALIHNTHLCFV 483
Db 361 FRNLSTFRNLFTIHGQQLMSMPALAIYVSSLSXLEMRNLKQISSGVVYIQHRDLCTV 420
Qy 484 HTVPMDQLFRNPQALLHTANRDEDC----- 510
Db 421 SNIRMPALIQEPEOKVWVENLRADLCGKFLITLISVQHNIIIMHIFAIKREKNNHLLGSV 480
Qy 511 ----- 510
Db 481 QRGRLLSWHSVSVLYQELQFQWHLHRLMLYIQVINSITQDSNEHQLTDACYSPSVPT 540
Qy 511 -----VG 512

```

```
Db 541 SUTIRARAYAIOSAGLAMELEBOTIARASASMRHSKITPAEGROVPRKMWELGVCASARAGIA 600
Oy 513 EGLA-----CHOLCARGHCWGPFTQCVCNSQFLRGQCEVEECVLOGLPREVY---N 562
Db 601 EPLARAVCRKCHPLCELCTNYGHEQVSKCHYRRRQCEEC-----PADHTDDE 654
Oy 563 ARHCLPCHPECCOPONGSVTCFPEADQVACAHYK-----DPPF-----CVARCPGG 609
Db 655 .ORECCORHPEC---NG---CTGPGADDCSKSCNFKLPDANETGPPYNSITMFNCTSKCPLE 708
Oy 610 VK-PDLSTWPIKFPDEEGACQPCPINCNSCDDLDKDCPAQGRASPLTSIVSAVGL 668
Db 709 MHWVYQTAIPY-----CAASPPRSKITANDL-----VNMFTITGAV 749
Oy 669 LVVVLGVVFGI-LIKRQOKIRRYT--MRRLQETELVEPLTPSGAMPQOAKRIKETE 725
Db 750 LVPTICILCVTVYICROKQAKKETVKMTALSGRDESEPLRPSNIGALCTRIYKDAE 809
Oy 726 LKRVKVLGSGAGFYTKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAVVMAGVS 785
Db 810 LRKGVLLGAGFGRVYKGVWPEGEVVKIPVAIKELKSTGAESEEFREAVIMASEEH 869
Oy 786 PVSRLIGICLTSTVOLVTLQMLPYGCLLDHVRNENRGSLSDLLMCMQIAGKSYLEBV 845
Db 870 VNLKLLAVCMSSQMLLITQLMPLGCLLDYVRNNRDKIGSKALLNMSTQIAGMSYLEEK 929
Oy 846 RLVRHDLAARNLVK---SPNHVKITDFGLARLLDIDETEHADGGKVPDKMALESILR 902
Db 930 RLVRHDLAARNLVVLLAGEH----DFGLAKLLSSDSMEYKAAGKMPKIMLALECIKN 985
Oy 903 RRFTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLIEKGBRLPQPICTIDVYMINV 962
Db 986 RVFTSKSDVMAGVIMWELTFGQRPHEMIPAKDIPDLIEVGLKLEQPEICSLDIYCTLL 1045
Oy 963 KCMWIDSECRPFRELVSEFSRMARDPQRFVVIQNEDELQ--PASPLDSFFYSLEDD-- 1018
Db 1046 SCHWIDAMRPFTKQTLTTFAPARDPGRYLAIIIGSKFTRLPA-----YTSQDKDLI 1098
Oy 1019 -DMGDLVDAEYLVPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEAR 1077
Db 1099 RKLAPTTDGSSEAIAPRDYLGKALGPS-----HRTDCT-----DENP 1137
Oy 1078 -----RSPLASSEGAGSVFPG---DLGMAKAGKIQSLPTHDPSPFLQYSDDPVLP 1128
Db 1138 KLMRYCKDPSNKSSTGDERDSSAREVGVNLR-----LDLPV 1176
Oy 1129 ETDGVAVLUTCSPOPEYVNPQVRPOPSPREGPLPAARAPAGATLERAKTISPKGNKVYK 1188
Db 1177 DEDDYLP-TCOPGRNNNNNNN-----NPNQNNMAAGVAAAGYM----- 1214
Oy 1189 DVFAFGAVENDEYL-----TPOGGAAPQH-----PPAFSP-AFDN 1225
Db 1215 DLIGVPSVDNPEYLLNMQTLGVGESPIPTQITGIPWCGPGTMEVKVYMPQSEPTSDH 1274
Oy 1226 LYIWD 1230
Db 1275 EYIWD 1279
```

RESULT 14

S35745
Protein-tyrosine kinase (EC 2.7.1.112) erdb - avian erythroblastosis virus
C/Species: avian erythroblastosis virus
C/Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C/Accession: S35745
R:Vennstrom, B.
Submitted to the EMBL Data Library, March 1993
A/Reference number: S35743
A/Accession: S35745
A/Molecule type: DNA
A/Residues: 1-544 <VEN>
A/Cross-references: EMBL:X12707

C/Genetics:
A/Gene: erdb
C/Superfamily: epidermal growth factor receptor; protein kinase homology
C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F:135-400/Domain: protein kinase homology <Kin>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #catalus Predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 8,6e-61;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

```
Oy 584 GPEADQVACAHYKDPFVVARCPGCVKDLSTYMPIKFPDEEGACQPCPINCSCVDL 643
Db 1 GP--DHCMCAHFIDQPHCVKACPAAGVLGENDTL-VWKADANAVALCHPNCSTRCKGP 57
Oy 644 DDKGCPAEGRASPLTSIVSAV-GILVYVGLVFPILIKRQOKIRKTMRLLOETEL 702
Db 58 GLEGP--NSKTPSIAGVVGGLCLVAVGIGLVYLR--HVRKTLRLLOREL 113
Oy 703 VEPLTPSGAMPQOAKRIKETEELRKVYLGSGAGFYTKGIWIPDGENVKIPVAIKVL 762
Db 114 VEPLTPSGAMPQOAKRIKETEELRKVYLGSGAGFYTKGIWIPDGENVKIPVAIKVL 762
Oy 763 ENTSPKANKEILDEAVVMAGVSPVYSRLGICLTSTVOLVTLQMLPYGCLLDHVRNENR 822
Db 174 EATSPKANKEILDEAVVMASVDPNPHVCRLLGICLTSTVOLVTLQMLPYGCLLDYIRHKN 233
Oy 822 LGSQDLNMCMQIAGKSYLEBVLRVHRLAARNLVKSPNHVKITDFGLARLLDIDETE 862
Db 234 IGSQYLLNMCMQIAGKSYLEBVLRVHRLAARNLVKTPQHVKITDFGLAKOLGADKE 293
Oy 883 YHADGKVPDKMALESILRRRTHOSDWSYGVTVWELMTFGAKPYDGI PAREIPDLIE 942
Db 294 YHAEKGKVPDKMALESILRRYTHOSDWSYGVTVWELMTFGSKRYDGI PASEISSVLE 353
Oy 943 KGERLPQPICTIDVYMINVCMWIDSECRPFRELVSEFSRMARDPQRFVVIQ-NEDELQ 1001
Db 354 KGERLPQPICTIDVYMINVCMWIDSECRPFRELVSEFSRMARDPQRFVVIQ-NEDELQ 1001
Oy 1002 PASPLDSFFYSLEDDGDLVDAEYLVPQGFPCPDPAAGAGVHHRRSSSTRSG 1061
Db 414 LPSPTDSKFRYTLMEEDMEDIVDAEYLVPQGF-----NSPST--- 454
Oy 1062 GGDLTGLLEPSEEARSPPL-----APSEGASDVVDGLGMAKAGKIQSLPTHDPSP 1116
Db 455 -----SRFLSSLSATSNKATNCLDRNG-----H----- 481
Oy 1117 RYSEDPVPLPSETDGYVAPLTCSPQPEYVNPQVRPOPSPREGPLPAARAPAGAT-LER 1175
Db 482 -----PYREDOFL-----PAPEYVNG--LMPKKRSTAMVQNGQYNYISLRAISK 523
Oy 1176 AKTLSPKGNKVYKDVFAFGAVENDEYL 1203
Db 524 LPIDSRYN-----SHSTAVDNPEYL 544
```

RESULT 15

S00727
Kinase-related transforming protein (erdb) (EC 2.7.1.-) - avian erythroblastosis viru
C/Species: avian erythroblastosis virus
C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C/Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A/Title: Common site of mutation in the erdb gene of avian erythroblastosis virus mut.
A/Reference number: S00727; MUID:88217326; PMID:2897102
A/Accession: S00727
A/Molecule type: DNA
A/Residues: 1-545 <SCO>
A/Cross-references: EMBL:X06943
C/Genetics:
A/Gene: erdb

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 24.0%; Score 1640; DB 2; Length 545;

Best Local Similarity 54.9%; Pred. No. 1.7e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

```
QY 584 GREADQVACAHYKDPFCVARGSPGKPDLSYMIKFPDEBGAQCPPTNCHSCVDL 643
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GP--DHCMKAHFIDGPHCVACGAVLGENDTL--VMKYADANAVCQLCHPCTGCKGP 57
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 644 DDKGCPAEORASPLTSIVSAV--GILLVVVLGVVFGLIKRQOKIRKYTRRLQETEL 702
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 GLECCP---NGSKTPSIAGVVGGLCLVVGGLGIGLYLR--HIVKRTIRLLQEREL 113
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 703 VEPLTPSGAMPNQAKMRLKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLR 762
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 VEPLTPSGEAPNQAHRLILKETEFKVKVILGFAGFTVYKGLWIPGEKVTIPVAIKELR 173
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 763 ENTSPKANKELDEAYVMAGSPYVSRILGICLTSTVQLVTOLMPYGCILDHYRENRGR 822
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 174 EATSPKANKELDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCILDHYIREHKN 233
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 823 LGSODLLNMCMQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 234 IGSQYLLNMCVQIAKGNVLEERLHVRLDLAARNVLKTPQDVKITDFGLAKQIGADEKE 293
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 883 YHADGGVPTIKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLE 942
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 YHAEGGVPTIKMALESILHRIYTHOSDWSYGVTVWELMTFGSKPYDGIIPASEISSYLE 353
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 943 KGERLPOPICTIVVMIMVCMWIDSECRPRELVESESRMARDPQRFVLIQ--NEDLG 1001
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 KGERLPOPICTIVVMIMVCMWIDSDSRPKPRELIAEFSKMARDPPEYLVLIQDERMR 413
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1002 PASPLDSTFYRSLEDDEMDGLVDAEELVYPOGFCPPDPAFGAGMVAHHRSSSTRSG 1061
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 414 LPSPTDSKFYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1062 GGDLLTGLPSEEEAPRSPL-----APSEGASDVFPDGLMGAKGLQSLPTHDPSPLO 1116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 -----SRTPLLSLSLATSNNSATNCIDRMNG-----H----- 481
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1117 RYSEDPVPLPSETDGVVAPLTCSPOPEYVNOQPVRRPOPSPREGPLPAARPAQAT--LER 1175
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 -----PYREDGFL-----PAPEYVNO--LMPKKPSTAMVONQIYVYISLTAISK 523
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1176 AKTLSPGKNGVAKDVFAFGAVENPEYL 1203
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 524 LPMDSRYON-----SHSTAVDNPEYL 544
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 22, 2003, 09:08:27
Job time : 32.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-103-117-14

Perfect score: 6847
Sequence: 1 MELALCRWGLLLALPPGA.....TFKGTPTAENPEYIGLDVIV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6708	98.0	1255	1	ERR2_HUMAN
2	5946	86.8	1257	1	ERR2_RAT
3	5921.5	86.5	1254	1	ERR2_MESAU
4	3137	45.8	1210	1	EGFR_HUMAN
5	3109	45.4	1210	1	EGFR_MOUSE
6	2986.5	43.6	1308	1	ERR4_HUMAN
7	2967	43.3	1308	1	ERR4_RAT
8	2695.5	39.4	1167	1	XMRK_XIPMA
9	2427.5	35.5	1342	1	ERR3_HUMAN
10	2357.5	34.4	1339	1	ERR3_MOUSE
11	1977	28.9	1426	1	EGFR_DROME
12	1749.5	25.6	634	1	ERRB_ALV
13	1703	24.9	604	1	ERRB_AVIER
14	1630	23.8	540	1	ERRB_AVIER
15	1594	23.3	703	1	EGFR_CHICK
16	1296	18.9	1323	1	L723_CAEEL
17	1142.5	16.7	245	1	ERR2_MOUSE
18	727	10.6	1363	1	ILPR_BRALA
19	706	10.3	1382	1	INSR_HUMAN
20	701	10.2	1382	1	INSR_RAT
21	700.5	10.2	1372	1	INSR_MOUSE
22	697	10.2	1300	1	INSR_MOUSE
23	687	10.0	1297	1	INSR_HUMAN
24	683.5	10.0	1300	1	INSR_MOUSE
25	681	9.9	1607	1	MLP7_LYMST
26	671	9.8	1477	1	HTK7_HYDAT
27	640	9.3	1367	1	IGIR_HUMAN
28	630	9.2	1373	1	IGIR_MOUSE
29	626.5	9.1	1370	1	IGIR_RAT
30	625	9.1	1390	1	INSR_AEDAE
31	616	8.9	2146	1	INSR_HUMAN
32	610	8.9	2146	1	INSR_DROME
33	597	8.7	987	1	EPB4_MOUSE

34	595.5	8.7	902	1	EPBB_XENLA	Q91736 xenopus lae
35	594	8.7	984	1	EPB1_RAT	P09759 ratius norv
36	592	8.6	985	1	EPBA_XENLA	Q91571 xenopus lae
37	591.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
38	589.5	8.6	977	1	EPB2_MOUSE	Q03145 mus musculu
39	588	8.6	984	1	EPB1_HUMAN	P54762 homo sapien
40	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
41	584.5	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
42	576	8.4	988	1	EPB3_CHICK	Q07498 gallus gall
43	575.5	8.4	998	1	EPB3_HUMAN	P54753 homo sapien
44	573.5	8.4	1053	1	FAK1_CHICK	Q09444 gallus gall
45	569	8.3	1068	1	FAK1_XENLA	Q91738 xenopus lae

ALIGNMENTS

```
RESULT 1
ERR2_HUMAN
ID ERR2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (rel. 05, Created)
DT 13-AUG-1987 (rel. 05, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; Pubmed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Sema K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor."
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; Pubmed=299974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Liberman T.A., Schlessinger J.;
RA Francke U., Levinson A., Ulrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene."
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; Pubmed=299567;
RA Sema K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RL human salivary gland adenocarcinoma."
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; Pubmed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization."
RL Genomics 15:426-429(1993).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
```


CC -1- PTH: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M11767; AAA35808.1; -
 DR EMBL: M11761; AAA35808.1; JOINED.
 DR EMBL: M11762; AAA35808.1; JOINED.
 DR EMBL: M11763; AAA35808.1; JOINED.
 DR EMBL: M11764; AAA35808.1; JOINED.
 DR EMBL: M11765; AAA35808.1; JOINED.
 DR EMBL: M11766; AAA35808.1; JOINED.
 DR EMBL: M11730; AAA35493.1; -
 DR EMBL: M12036; AAA35978.1; -
 DR EMBL: X03363; CAA27060.1; -
 DR PIR: A25491; A25491.
 DR PIR: A24571; A24571.
 DR HSSP: P11352; 1FGK.
 DR Genem; HGNC:3430; ERBB2.
 DR MIM; 164870; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_mocif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR Prodom: P000001; Euk_kinase; 1.
 DR SMART; SM00261; Fu; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRAASMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 I -> V.
 FT VARIANT 655 655 /FTID=VAR_004077.
 FT VARIANT 655 655 I -> V.
 FT CONFLICT 1170 1170 /FTID=VAR_004078.
 FT CONFLICT 1255 AA; 137909 MM; 39E9DFDA04DCF962 CRC64;
 SQ SEQUENCE

Query Match 98.0%; Score 6708; DB 1; Length 1255;
 Best Local Similarity 98.1%; Pred. No. 0;
 Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELAALCRWGLLALLPRGAASQVCTGDMKRLRLASPEPTHLDMLRHLYOGCOVQGNL 60
 DB 1 MELAALCRWGLLALLPRGAASQVCTGDMKRLRLASPEPTHLDMLRHLYOGCOVQGNL 60
 QY ELTYLPTNALSFLQDIOGVGVLIJAHNQVROVPRQRLRYVGTQLFEDNYALAVLDNG 120
 DB ELTYLPTNALSFLQDIOGVGVLIJAHNQVROVPRQRLRYVGTQLFEDNYALAVLDNG 120
 QY 61 ELTYLPTNALSFLQDIOGVGVLIJAHNQVROVPRQRLRYVGTQLFEDNYALAVLDNG 120
 DB 61 ELTYLPTNALSFLQDIOGVGVLIJAHNQVROVPRQRLRYVGTQLFEDNYALAVLDNG 120
 QY 121 DELNENFTVSFWLKVPRVAS---HLEQLRLTEILKGVLIQRPOLCYQDTILMK 176
 DB 121 DELNENFTVSFWLKVPRVAS---HLEQLRLTEILKGVLIQRPOLCYQDTILMK 176
 QY 121 DELNENFTVSFWLKVPRVAS---HLEQLRLTEILKGVLIQRPOLCYQDTILMK 176
 DB 121 DELNENFTVSFWLKVPRVAS---HLEQLRLTEILKGVLIQRPOLCYQDTILMK 176
 QY 177 DI FHKNOALTLIDNBRACHPCSPMKSGSCWSESSDDCSLRTVACGACACCKP 236
 DB 177 DI FHKNOALTLIDNBRACHPCSPMKSGSCWSESSDDCSLRTVACGACACCKP 236
 QY 171 DIFHKNOALTLIDNBRACHPCSPMKSGSCWSESSDDCSLRTVACGACACCKP 230
 DB 171 DIFHKNOALTLIDNBRACHPCSPMKSGSCWSESSDDCSLRTVACGACACCKP 230
 QY 237 LPTDCHEGCAAGCTGPKASDCLAHFNHSGICEHPCALVYNTDFESMNPGRYT 296
 DB 237 LPTDCHEGCAAGCTGPKASDCLAHFNHSGICEHPCALVYNTDFESMNPGRYT 296
 QY 231 LPTDCHEGCAAGCTGPKASDCLAHFNHSGICEHPCALVYNTDFESMNPGRYT 290
 DB 231 LPTDCHEGCAAGCTGPKASDCLAHFNHSGICEHPCALVYNTDFESMNPGRYT 290
 QY 297 FGASCTACPYNYLSTDVGSCTLVCPAHNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 356
 DB 297 FGASCTACPYNYLSTDVGSCTLVCPAHNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 356
 QY 291 FGASCTACPYNYLSTDVGSCTLVCPAHNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 350
 DB 291 FGASCTACPYNYLSTDVGSCTLVCPAHNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 350
 QY 357 REVRANTSNIOEFAGCKKIFGSLARLPSSFGDDPASNTAPLOPELOVPELLEITGYL 416
 DB 357 REVRANTSNIOEFAGCKKIFGSLARLPSSFGDDPASNTAPLOPELOVPELLEITGYL 416
 QY 351 REVRANTSNIOEFAGCKKIFGSLARLPSSFGDDPASNTAPLOPELOVPELLEITGYL 410
 DB 351 REVRANTSNIOEFAGCKKIFGSLARLPSSFGDDPASNTAPLOPELOVPELLEITGYL 410
 QY 417 YISAMPDSLPLDSVFNQLOVIRGRILHNGAYSLLTQGLGISWLGRLSRELSSGLALIH 476
 DB 417 YISAMPDSLPLDSVFNQLOVIRGRILHNGAYSLLTQGLGISWLGRLSRELSSGLALIH 476
 QY 411 YISAMPDSLPLDSVFNQLOVIRGRILHNGAYSLLTQGLGISWLGRLSRELSSGLALIH 470
 DB 411 YISAMPDSLPLDSVFNQLOVIRGRILHNGAYSLLTQGLGISWLGRLSRELSSGLALIH 470
 QY 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECEGEGLAHQLCARGHGWGPQCVN 536
 DB 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECEGEGLAHQLCARGHGWGPQCVN 536
 QY 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECEGEGLAHQLCARGHGWGPQCVN 530
 DB 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECEGEGLAHQLCARGHGWGPQCVN 530
 QY 537 CSQPLRGQCEVECEVLOGLPREYVNAARCLCHPECCQONGSYTCFGEADQVCAH 596
 DB 537 CSQPLRGQCEVECEVLOGLPREYVNAARCLCHPECCQONGSYTCFGEADQVCAH 596
 QY 531 CSQPLRGQCEVECEVLOGLPREYVNAARCLCHPECCQONGSYTCFGEADQVCAH 590
 DB 531 CSQPLRGQCEVECEVLOGLPREYVNAARCLCHPECCQONGSYTCFGEADQVCAH 590
 QY 597 KQPPCVARCPGAVPDLSSYPMIKFPEDEGACQCPINCTHSCVDLDDKGCFAEORASP 656
 DB 597 KQPPCVARCPGAVPDLSSYPMIKFPEDEGACQCPINCTHSCVDLDDKGCFAEORASP 656
 QY 591 KQPPCVARCPGAVPDLSSYPMIKFPEDEGACQCPINCTHSCVDLDDKGCFAEORASP 650
 DB 591 KQPPCVARCPGAVPDLSSYPMIKFPEDEGACQCPINCTHSCVDLDDKGCFAEORASP 650
 QY 657 LTSIYSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEBLTPSGAMPND 716
 DB 657 LTSIYSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEBLTPSGAMPND 716
 QY 651 LTSIYSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEBLTPSGAMPND 710
 DB 651 LTSIYSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLQETELVEBLTPSGAMPND 710

```

OY 717 QMRILKETELRKVKVLCGACGCTVYKGIWIPDGENVKIPVAIKVIRENTSPANKELIDE 776
Db 711 QMRILKETELRKVKVLCGACGCTVYKGIWIPDGENVKIPVAIKVIRENTSPANKELIDE 770
OY 777 AYVWAGSPYVSRLLGICLTSTVOLVQOLMPYGCILDHVRNRRGLSGODLLNMCQIA 836
Db 771 AYVWAGSPYVSRLLGICLTSTVOLVQOLMPYGCILDHVRNRRGLSGODLLNMCQIA 830
OY 837 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEYHADGKVPVKMA 896
Db 831 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEYHADGKVPVKMA 890
OY 897 LESILRRRFTHOSDVWSYGVTVWEIMLTFCAPKYDGIIPAREIDPLEKGRLLPQPICTID 956
Db 891 LESILRRRFTHOSDVWSYGVTVWEIMLTFCAPKYDGIIPAREIDPLEKGRLLPQPICTID 950
OY 957 VYMIWVKCMITDSECRPRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYMIWVKCMITDSECRPRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
OY 1017 DDDMGDLVDAEEYLVPOQGFPCPDPAFGAGVHHRHSSSTRSGGDLTTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEEYLVPOQGFPCPDPAFGAGVHHRHSSSTRSGGDLTTLGLEPSEEA 1070
OY 1077 PRSPLAPSEAGSDVFDGDLGMAAKGLQSLFTHDPSLQRYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEAGSDVFDGDLGMAAKGLQSLFTHDPSLQRYSEDPVPLPSETDGYVAP 1130
OY 1137 LTCSPOPEYVNOPDVRPQPSREGLPLAPAPAGATLEAKTKLSGKNGVWVDVAFGA 1196
Db 1131 LTCSPOPEYVNOPDVRPQPSREGLPLAPAPAGATLEAKTKLSGKNGVWVDVAFGA 1190
OY 1197 VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDODPPERGAPESTFKGTPTAENPEYL 1256
Db 1191 VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDODPPERGAPESTFKGTPTAENPEYL 1250
OY 1257 LDVPV 1261
Db 1251 LDVPV 1255

```

```

RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofis F.J., Doak D.G., Mulvey D.,
RT Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.,
RT "Three dimensional structure of the transmembrane region of the proto-
RL oncogene and oncogenic forms of the new protein."
CC EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEURULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEURULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR PIR, X03362; CAA27059.1; ALT_INIT.
DR EMBL, A24562; TVRTNU.
DR HSSP, P13362; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002145; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART, SM00261; FU; 3.
DR SMART, SM00219; TyKc; 1.
DR PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE, PS00111; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1257
FT DOMAIN 22 654
FT TRANSMEM 655 677
FT DOMAIN 678 1257
FT DOMAIN 159 369
FT DOMAIN 473 646
FT DOMAIN 722 989
FT NP_BIND 728 736
FT BINDING 755 755
FT ACT_SITE 847 847
FT DISULFID 196 205
FT DISULFID 200 213
FT DISULFID 221 228
FT DISULFID 225 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 312
FT DISULFID 316 332
FT DISULFID 335 339
FT DISULFID 513 522

```

FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.
 FT DISULFID 569 585 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA, 138831 MW, 6129264583011402 CRC64.

Query Match 86.8%; Score 5946; DB 1; Length 1257;
 Best Local Similarity 87.1%; Pred. No. 1,8e-311;
 Matches 1099; Conservative 50; Mismatches 107; Indels 6; Gaps 3;

QY 1 MELALCRWGLLALIPGAASTQVCTGIDMKLRLEPASPTHLDMRLHYQGQVVGNL 60
 DB 1 MELAACRWGFLALLPFGIAGTQVCTGIDMKLRLEPASPTHLDMRLHYQGQVVGNL 60
 QY 61 ELTYLTNASTSLFDIOEYGVVLLAHNOVROVRLRIYVGTQLFENYALAVLDNG 120
 DB 61 ELTYLPANASLSLFDIOEYGVVLLAHNOVKRVLQRLRIYVGTQLFEDKYALAVLDNR 120
 QY 121 DPLANNFTVFSFWLVPKVASHLEQLSLTEILKGVLLIQNNPOLCYODTILMKDIFH 180
 DB 121 DDPDNNVAASRPG--RTP--GLRELQRLSLTEILKGVLLIRGNPOLCYODMLMVDVER 175
 QY 181 KNNQALLTIDTRSRACHPSCMKGSRGWSSSEDCQSLRTVACGACGKGLPTD 240
 DB 176 KNNQALPVIDITRSRACPCAPACDNHMGWSPBCQILTGITCTSGACGKGLPTD 235
 QY 241 CCHGOCAGCTGPKHSDCLAFHNHSGICELHCPALVTNTDTPFSPMPREGRTFGAS 300
 DB 236 CCHGOCAGCTGPKHSDCLAFHNHSGICELHCPALVTNTDTPFSPMPREGRTFGAS 295
 QY 301 CVTACPYNYLSTDVSGCTIIVCPPLHNOEVTAEADGTORCEKSKPCARVYCYLGMEHLREYR 360
 DB 296 CVTTCYNYLSTEVSGCTIIVCPPNNOEVTAEADGTORCEKSKPCARVYCYLGMEHLREYR 355
 QY 361 AATSAIOEPAGCKRTFGSLAFIPESFDGPASNTAPLOEQLQVETLEITGYLYISA 420
 DB 356 AATSDVVOEFDDCKRTFGSLAFIPESFDGPSSGIALPRLPEQLQVETLEITGYLYISA 415
 QY 421 WPDSLPDLGVFQNLQVIRGRIILHNGAYSLTLOGLISWLGRLSRLREISGLALIHNTTH 480
 DB 416 WPDSLPDLGVFQNLRIIRGRILHDGAYSLTLOGLISWLGRLSRLREISGLALIHNTTH 475
 QY 481 CFVHTVPMDQLFENPHQALLHTANRPDE- CVGEGLACHOLCARGHCKWGPPTQCVNCSQ 539
 DB 476 CFVHTVPMDQLFENPHQALLHSGNRPEDLCVSSGLVCNSLCHNGHCKWGPPTQCVNCSH 535
 QY 540 FTRGOCVECRVLOQLPREYVNAKRLCPHPECQONQNSVTFEGEADQCVACHYKDP 599
 DB 536 FTRGOCVECRVLOQLPREYVNAKRLCPHPECQONQNSVTFEGEADQCVACHYKDS 595
 QY 600 PCVVARCPGSKPDLSTYMPIMKFPDEGACQPCPINCTSHCVLDLKGCAEORASPLTS 659
 DB 596 SSCVVARCPGSKPDLSTYMPIMKFPDEGACQPCPINCTSHCVLDLKGCAEORASPLTS 655
 QY 660 IYSAVVGILLVVLGVVFGILLKRRQOKIRKYMRLLOETELVEPLTSGAMPNOQNR 719
 DB 656 IYATVVGILLVVLGVVFGILLKRRQOKIRKYMRLLOETELVEPLTSGAMPNOQNR 715
 QY 720 ILKETELRKVKVLGSAFGTVVYGIWIPGGENKIVALKVLRENTSPPANKELIDEAYV 779

DB 716 ILKETELRKVKVLGSAFGTVVYGIWIPGGENKIVALKVLRENTSPPANKELIDEAYV 775
 QY 780 MAGVSPYVSRLLGICLTSTVOLVTLQMPGCLLDVRENRGLSGODLNMCMQIAKGM 839
 DB 776 MAGVSPYVSRLLGICLTSTVOLVTLQMPGCLLDVRENRGLSGODLNMCMQIAKGM 835
 QY 840 SYLEDRLVHRDLAANVLYKSPNHKRTDPGLARLLDIDETETHADGCKVPIKMMALLES 899
 DB 836 SYLEDRLVHRDLAANVLYKSPNHKRTDPGLARLLDIDETETHADGCKVPIKMMALLES 895
 QY 900 ILRRRFTHSDVSYCVTWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYM 959
 DB 896 ILRRRFTHSDVSYCVTWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYM 955
 QY 960 IMVKCMIMSECCPRRELVSERSRAARPORVVIQNEGLGASPLDSTFFYSLLLEDD 1019
 DB 956 IMVKCMIMSECCPRRELVSERSRAARPORVVIQNEGLGASPLDSTFFYSLLLEDD 1015
 QY 1020 MGDLYDAEYLVPOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079
 DB 1016 MGDLYDAEYLVPOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1075
 QY 1080 PLAPSEGAGSDVFDGDLGMAAKGLSLPTHDPSPLQRYSEDEPTVLPSETDGYVAPLTC 1139
 DB 1076 PLAPSEGAGSDVFDGDLGMAAKGLSLPTHDPSPLQRYSEDEPTVLPSETDGYVAPLTC 1135
 QY 1140 SPOPEYVNPDPVRPQPPSREGPLPAARAGATLEBAKTLSPKNGVYKDVAFGAVEN 1199
 DB 1136 SPOPEYVNPDPVRPQPPSREGPLPAARAGATLEBAKTLSPKNGVYKDVAFGAVEN 1195
 QY 1200 PEYLTPOGGAAPQPPPPAFSPAFDNLVYWDOPPERPAPSPSTFFKGPPTAENEYGLDV 1259
 DB 1196 PEYLTPOGGAAPQPPPPAFSPAFDNLVYWDOPPERPAPSPSTFFKGPPTAENEYGLDV 1255
 QY 1260 PV 1261
 DB 1256 PV 1257

RESULT 3
 ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC 060553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (neu proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN 1) _SEQUENCE FROM N.A.
 RP MEDLINE=94193007; PubMed=7908275;
 RC TISSUE=Nerve;
 RX Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
 RL Gene 140:251-255(1994).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

CC RESIDUES.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D16295; BAA03801.1; -
 CC HSSP: P11362; 1FCX.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU_3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1254
 FT DOMAIN 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1254
 FT DOMAIN 158 358
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT CARBOHYD 658 658
 FT VARIANT 659 659
 FT SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 86.5%; Score 5921.5; DB 1; Length 1254;
 Best Local Similarity 86.7%; Pred. No. 3,6e-310;
 Matches 1093; Conservative 58; Mismatches 103; Indels 7; Gaps 3;
 QY 1 MELALCRGGLLALLPPAASTOVCTGDMKRLPASETHLDMRLHYOGCOVQGNL 60
 DB 1 MELAMCGGLLALLSPASGVCTGDMKRLPASETHLDIRHHYOGCOVQGNL 60
 QY 61 ELTYLPTNASLFLDIOEVGVYLIANQVAVPQRLRIYRGTOLEFENVYLAVALDNG 120
 DB 61 ELTYLPANATLSFLDIOEVGYMLIAHQVAVPQRLRIYRGTOLEFENVYLAVALDNR 120
 QY 121 DELANNFNFTVSAFWLVPKVASHLLEQLSLTEILKGVYLIQRNPOLCYODTLIMKDFH 180
 DB 121 DELDNNVTATG---RTPE--GLRELQLASLLEILKGVYLIQRNPOLCYODTLIMKDFH 174
 QY 181 KNNQALATLIDNRSRACHPCSPCKSGASPCWBSSEDCOSLRTVCAGCCARCKGPLPTD 240
 DB 175 KNNQALAPVDIDNRSRACPCPACACKDNHCWASPEDCOTLICTIAPRAVAPARARLPTD 234
 QY 241 CCEBOCAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDFESMNPREGRTYFGAS 300
 DB 235 CCEBOCAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDFESMNPREGRTYFGAS 294
 QY 301 CYTACPPYNYLSTDVGSCTLVCPDLHNOEYVAEDGTORCEKSKPCARVCYGLGMEHLREVR 360
 DB 295 CYTTCPPYNYLSTEVGSCTLVCPDLNNOEYVAEDGTORCEKSKSCARVCYGLGMEHLRGAR 354
 QY 361 AVTSANIOEFACCKKIFGSLAFLESFDDGPASNTAPLOPEOLOVETLEITGYLYISA 420
 DB 355 AVTSANIOEFACCKKIFGSLAFLESFDDGPASNTAPLOPEOLOVETLEITGYLYISA 414
 QY 421 WPDLSLDLSVFNQLQVIRIRILIANCAVSLTLOGIGISWGLASRLSGGLAIHNNTHL 480
 DB 415 WPDLSLDLSVFNQLQVIRIRILIANCAVSLTLOGIGISWGLASRLSGGLAIHNNTHL 474
 QY 481 CFVHTVPMQDLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWCGPPTCVNCSOF 540
 DB 475 CFVHTVPMQDLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWCGPPTCVNCSOF 534
 QY 541 LRQECVEBCRYLQGLPREYVNAHICLPCHPECPQNGSVTCFGEADOCVACAHYKDP 600
 DB 535 LRQECVEBCRYLQGLPREYVNAHICLPCHPECPQNGSVTCFGEADOCVACAHYKDP 594
 QY 601 FCVARCPGVKXDSLTPMIPKRPDEEGACQPCINCTHSVDLDDKGCAGAEQASLTSI 660
 DB 595 FCVARCPGVKXDSLTPMIPKRPDEEGACQPCINCTHSVDLDDKGCAGAEQASLTSI 654
 QY 661 VSAVVGILLVVLGVVFGILIKRPOOKIRKYMTRILQETELVEPLTPSGAMPNOAKMI 720
 DB 655 IATVVGILLVVLGVVFGILIKRPOOKIRKYMTRILQETELVEPLTPSGAMPNOAKMI 714
 QY 721 LKETELRKVKVLSGAFGVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYVM 780
 DB 715 LKETELRKVKVLSGAFGVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYVM 774
 QY 781 AGVSPYYSRLIGTLSTVQLVTLQMPYGCLLDHYREKRGVLSGODLIMCMQIAKGS 840
 DB 775 AGVSPYYSRLIGTLSTVQLVTLQMPYGCLLDHYREKRGVLSGODLIMCMQIAKGS 834
 QY 841 YLEDRVLVHRDLAARNVLYKSPNHKITDFGLARLLDIDETRYHADGGKVP1KMMALESI 900
 DB 835 YLEDRVLVHRDLAARNVLYKSPNHKITDFGLARLLDIDETRYHADGGKVP1KMMALESI 894
 QY 901 LRRRFTHOSDVSYGVTWVMEIMTFGAKPYDGIIPAREIDPLEKGRLLPQPICTIDVYMI 960
 DB 895 LRRRFTHOSDVSYGVTWVMEIMTFGAKPYDGIIPAREIDPLEKGRLLPQPICTIDVYMI 954
 QY 961 MVKCMIDSECRPRPRELVSFESRVARDPORFVYIQNEDLGPASLSTFTRSLLEDDDM 1020
 DB 955 MVKCMIDSECRPRPRELVSFESRVARDPORFVYIQNEDLGPASLSTFTRSLLEDDDM 1014

QY 1021 GDLVDAEEYLVQOQGFCEPDPAPGAGGVVHHRRSSSTSSGGGDLTLGLPSEEEANRSP 1080
DB 1015 GDLVDAEEYLVQOQGFCEPDPAPGAGSTAHRRSSSTSSGGGDLTLGLMGPCEBPPRSP 1074
QY 1081 LAPSEGAGDVDPDGLGMAAGLQSLPHHDSPLQRYSEDPRVPLPSTETDGVAPLQTS 1140
DB 1075 LAPSEGAGDVDPDGLGMAAGLQSLPHHDSPLQRYSEDPRVPLPSTETDGVAPLQTS 1134
QY 1141 POPEYVNOPDVRPOPSPREGPLPAARPAATLERAKTSLSPGKNGVVKDVFARFGAVENP 1200
DB 1135 POPEYVNOPDVRPOPSPREGPLPAARPAATLERAKTSLSPGKNGVVKDVFARFGAVENP 1194
QY 1201 EYLTPOGGAAPQHPHPAPSPAFNDLYWDODPREGAPSPSTFKGPTAENPEYGLDVP 1260
DB 1195 EYLTPOGGAAPQHPHPAPSPAFNDLYWDODPREGAPSPSTFKGPTAENPEYGLDVP 1253
QY 1261 V 1261
DB 1254 V 1254

RESULT 4
EGFR_HUMAN STANDARD, PRT, 1210 AA.
ID EGFR_HUMAN
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayer E.L.V., Whitte N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654366;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN (3)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maibhe N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN (4)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Garfili J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN (5)
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel Sinclair C., Pearseall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN (6)
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielson A.J., Scheel C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibhe N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN (7)
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krutiger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN (8)
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAe overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN (9)
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN (10)
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whitte N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN (11)
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN (12)
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN (13)
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN (14)
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Wroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor


```

QY 963 KCMWIDSECRPRELWSEFSRMAPDQRFVIO-NEDLGPASPLDSTFYRSLLEDMDG 1021
DB 949 KCMWIDADSRKRFRELIERFSKMAKDQRLVIGQDGRMLPSPFTDSNFRALMDEEDMD 1008
QY 1022 DLVDAEYLVPOCGFFCPDPAAGAGMHHRRSSSTRSGGDLTGLFSEEBAPSP 1081
DB 1009 DVIDADEYLVPOCGFF-----SSPSTSTRPL 1034
QY 1082 APESGSDVFDLGMGAAGLQSLPTHPSPLORESPTVPLPSET--DGVAVLTC 1139
DB 1035 LSSLSTSN--NSTVACIDNGLOSCPIKEDSFLOKISSDPTGALTEDSIDDTFL----- 1087
QY 1140 SPOPEYVNOPDVAPPPSPRECPPLPAAPAGATLERAKTLPGRNGVVKDVFAGAVEN 1199
DB 1088 -PVPEYINQ-SVPRKPAQSVQNPVYHNQPLNP-----APSRDPHYQD--PHSTAVGN 1135
QY 1200 PERYL-TPQGAAPQPPPPPAFPFNDLYWDQ-----DP-----PERGAPST 1242
DB 1136 PEYLVNTVQ-----PTCVNSTFDSPAHMAQKSHQISLDMPDYQODFFPKAKENG 1186
QY 1243 FKCTPAENPEYL 1255
DB 1187 FKGS-TAENAYL 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
ID EGFR_MOUSE
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=93026370; Pubmed=1408137;
RA Avivi A., Skorecki K., Yavon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; Pubmed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; Pubmed=8125255;
RA Lueteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; Pubmed=2030916;
RA Avivi A., Tax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

```

```

RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X78987; CA55587.1; -
DR EMBL: U03425; AAA17899.1; -
DR EMBL: X59698; CAA42219.1; -
DR EMBL: L06864; AAA53029.1; -
DR EMBL: Z12608; CAA78249.1; -
DR HSSP: P11362; 1FGK.
DR WGD; MG1:95294; Egfr.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647
FT TRANSMEM 648 670
FT DOMAIN 671 1210
FT REPEAT 75 300
FT REPEAT 390 600
FT DOMAIN 1028 1071
FT DOMAIN 714 981
FT NP_BIND 720 728
FT BINDING 747 747
FT ACT_SITE 839 839
FT DISULFID 190 199
FT DISULFID 194 207
FT DISULFID 215 223
FT DISULFID 219 231
FT DISULFID 232 240
FT DISULFID 236 248
FT DISULFID 251 260
FT DISULFID 264 291
FT DISULFID 295 307
FT DISULFID 311 326
FT DISULFID 329 333

```



```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46D2D2F5 CRC64;

Query Match 45.4%; Score 3109; DB 1; Length 1210;
Best Local Similarity 49.1%; Pred. No. 2,1e-159;
Matches 629; Conservative 171; Mismatches 360; Indels 120; Gaps 24;

QY 11 LLLALLPFGAA--STQVCTGTDMLRLPASPTHLMDLRLHLYQGCGVVOGNETLYLPTN 68
DB 14 LLLTLLCAAGALAEKVKCOGTSNRLTOLGTFEDHFLSLQRMVNCEVVLGNLEITYVORN 73
QY 69 ASLSFLQDIOEVGVYLLIAHNOVROVRLRIYVGTOLFEDNYALAVLDNDGPLMNFNN 128
DB 74 YDLSFLKTIQEVAGVLIATLNTVERIPLENLOIIGNALYENTVALATLSNG----- 126
QY 129 FTVSFWLVRPKVSAHLEOL--RSLTLEILKGVLIOBNPOLCYOPTILMKDI----FKN 182
DB 127 -----TNRGLRELPMRNLOEILIGAVFRSNPILCMOTLIQWRDLVQVFNMSN 175
QY 183 NQLALTLIDTKRSRACHCSPMKCSRCWGESSEPCQSLTRTVCAAGCA-RCKGPLPTDC 241
DB 176 MSMDL---QSHPSSCPCKDPSCPNGSCMGGEENCOKLTIKICAQOCSHRCGRSPSDC 231
QY 242 CHEGCAAGCTGPKHSDCLAGLHFNHSGICELHCPALVYNYDFTFSMNPREGYTFGASC 301
DB 232 CHNOCAAGCTGPRSDCLVCOCKFODEATCKOTCPRLMYNPTTYOMDVNBEKYSFGATC 291
QY 302 VTACPYNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKSCRCARVAVCYGLGMEHLREYRA 361
DB 292 VKKCPRNVTVDHSGCVACGPDYEV-EDDGIRCKCKCDGCRVAVCNGIGIGERKDLIS 350
QY 362 VTSANIQEFAGCKKIFGLSLALPBEFDPPASNTAPLOPEOLQVETLEITGYLYISAM 421
DB 351 INANIKRKFYKCTAISGDLHLTPVAFKGDSTFRTPPLPRELEIKTYKEITGFLLIQAW 410
QY 422 PDSLPLDLSVFONLQVIRGRIILHNGAVSLTLOGLISWLTGLSLREISGLALIHNTILC 481
DB 411 PDNMTDLHAFENLEIRGTRKQHGQFSLAVVGLNTISGLSLSKISGDAVYIISGNRLVC 470
QY 482 FVHTVPMDOLEFRNPHQALLHTANREDECEVEGLACHQLCARHGCWGPPTQCVNCSQFL 541
DB 471 YANTINWKLFGCTPNQCKTKIMNNAEKCKAVNHNVCNPLCSSEGCGWGPPEPDCVSCQVNS 530
QY 542 RGGQVECEGRVLQGLPREYVNAHRLCPHRECQFQNGSVTCRGPBADCCVACAHYKDPDF 601

```

```

DB 531 RRECEKKNILGEPREFEVENSECICQCHPECLPQAMNITTCGRGPNDCIQCHAYIDGP 590
QY 602 CVARCPGKPDLSWPIKFPDEEGACCPCPINCHSCVDLDDKCCPAEQRASPLTSIV 661
DB 591 CYKTPAGIMGNNTL-VKTYADANNVCHLCHANCTYGCAGGELQCEWPPSPKIPSLA 649
QY 662 SAIVGILLVVLGVVFGI-LIKRROOKIRKTYMRRLQETELVEPLTPSGAMPNOAMRI 720
DB 650 TQIVGGLFLIV-VALGIGLFMRRIHYKRTLRLLQRELVEPLTPSGEANQHLRI 708
QY 721 LKETELRKVKYLGSGAFGVYVYGVIMPDGENVKIPALIVLAEKNSPKANKELIDAVYM 780
DB 709 LKETEFKIKIYVLSGSAFGVYVYKGLWIPSEGEKYPVAILKELBEATSPKANKETLDEAYM 768
QY 781 AGVSPVYSRLIGICLSTVOLTOLMPYGLLDHRENRGLGSGDLDLNMWQIAKGS 840
DB 769 ASVDNPHVCRLLGICLSTVOLTOLMPYGLLDHRENRGLGSGDLDLNMWQIAKGMN 828
QY 841 YLEDVRLVHRDLAANNVLKSPNHVKTDFGLARLIDIDETEHADGKVPKIMMALESI 900
DB 829 YLEDRLVHRDLAANNVLKTPQHYKITDFGLAKLLGAEKKEYHAEKGVPIKMALESI 888
QY 901 LRRRTTQSDVMSYGVYTWELMTFGAKPYDGIPIABEIPDLLEKGERLPQPICTIDVMI 960
DB 889 LHRITTHOSDVMSYGVYTWELMTFGSKPYDGIPIASDISILEKGERLPQPICTIDVMI 948
QY 961 MWKCMIMISCECPRERELVESEFRRAPDQRPVVQ-NEDDLGPASPLDSTFTRSLLEDD 1019
DB 949 MWKCMIMIDSPKRELLEFSSKMARDPQRYLVIOGDHRLMPSPTDSNFYRALMDEED 1008
QY 1020 MGDLVDAEYLVLPQOGFCPPDPAPAGGVNHRHSSSTRSGGDLTLLEPSEEEAPPS 1079
DB 1009 MEDVDADAEYLIPQGGF-----NSPST-----SRT 1034
QY 1080 PLAPSEAGSDVFDGLGKAAGLQSLFTHDPSLPQRYSEDPYPLPSET--DGVAAPL 1137
DB 1035 PLSSLSATSN-----NSTVACINRNGSCVKEDAFQRYSDPTGAVTEINDIDATL--- 1087
QY 1138 TCSPPQEVYNODVAPPPSPREGPLPAARPAATLERAKTISPGKNGVVKVDFAFGAV 1197
DB 1088 ---PVPEYVNO-SVEXRPAQSVQNPVYHNPQLHP-----APGDLHYQN--PHSNAY 1133
QY 1198 ENPEYL-TPQGAQAQPPPPAFSPAFDLVYWDQ-----DP-----PERGAP 1240
DB 1134 GNPEYLNTAQ-----PTCLSSGFSNPSALMIQKSHQMSLDNPDYQODFFPKTKPR 1184
QY 1241 STFKGTPAENEYGLDVP 1260
DB 1185 GIFKG-PTAENAEYLRAVP 1203

RESULT 6
ERB4_HUMAN STANDARD; PRT; 1308 AA.
ID ERB4_HUMAN
AC 015303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-tye cell surface receptor HER4).
GN ERBB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma.
RX MEDLINE=93189574; PubMed=8383326;
RA Ploman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.,
RA "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

```

RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC TISSUE=Fetal brain; PubMed=9334263;
RA Elenius K., Cortes G., Paul S., Choi C.J., Rio C., Plozman G.D.,
Klagsbrun M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester.";
RL J. Biol. Chem. 272:26761-26768 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN. HEART,
CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC LUNG, SALIVARY GLAND, AND PANCREAS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L07868; AAB59446.1; -;
DR HSSP; P11362; IFGK.
DR Genew; HGNC:3432; ERBB4.
DR MIM; 600543; -;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SMO0261; FU; 4.
DR SMART; SMO0219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.

FT	NP_BIND	724	732	ATP (BY SIMILARITY).
FT	ACT_SITE	751	751	ATP (BY SIMILARITY).
FT	ACT_SITE	843	843	BY SIMILARITY.
FT	DISULFID	189	197	BY SIMILARITY.
FT	DISULFID	193	205	BY SIMILARITY.
FT	DISULFID	213	221	BY SIMILARITY.
FT	DISULFID	217	229	BY SIMILARITY.
FT	DISULFID	230	238	BY SIMILARITY.
FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPTSHDCITTYWMTGSHLPQNA -> IGSSIEDCIGLMD (IN ISOFORM JM-B).

SEQUENCE 1308 AA; 146807 MW; 5E4AE8098508761 CRC64;
Query Match 43.6%; Score 2986.5; DB 1; Length 1308;
Best Local Similarity 45.2%; Pred. No. 8.5e-153;
Matches 612; Conservative 185; Mismatches 374; Indels 183; Gaps 29;
9 WGLLALLPPGAA---STOVCTGTDMLRLPASPEHLLDMLRLHYOGCOVQGNILELY 64
8 WWWSLVAAGTVQPSDSQSVACGTENKLSISLDEQYBALRKYYENCEVVMGNLEITS 67
65 LPTNASLSFLDIOGVGVLIANHVOVPLQRLRIYRGTOFLFEDNYALAVLDNGDPLN 124
68 IEHNNDLSFPLRSVREVTYVVALNQFRYLLPLENRIIRGTLYEDRYALALFLVYRKDG 127
125 NFNNFTVSFMLRVPKVSASHLEO--LRSTELIKGVGLIORPOLCYOPTILMKDIFKKN 182
128 NFG-----LQELGKXLTLELNGSVYVDQKFLCYADTIHMOIVRNRP 170
183 NQALATLIDTRNSRACHPCSPCKSGRCWGSESSDQSLTRTVACGC-ARCKGPLPTDC 241
171 WPSNLTLVSTNGSSGCRCHKSCGS-RCMGPTENHCQTLTRTVACAEQDGRCYGVYSDC 229
242 CHEQCAACCTCPKHSDCIACHFNHSGICELHCPALVYVNDTESMPNPSRRTFFGASC 301
230 CHRECAAGCGSPKQTDPCACNPFNDSGACVYQCQPTFFVYNPTTFPLEHNFNAKYTYGAF 289
302 VTACPYNYLSTDVGSCTLVCPFLHNOEYTAEDQTCERKSGKRCARVCYGLGMEHLREVRA 361
290 VKKCPHNIV-VDSSCAVACPSRSKREV-ENNGIKKCKCTDLCPCACGIGTGLSMASQT 347
362 VTSANIOEPAQCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEBITGYLYISAW 421

```

Db 348 VDSNDKFNCTKINGNLIFLVYGIHGPVNAIEAIDPEKLNVFRTVEITGFLNIQSK 407
Qy 422 PPSLDLSVFOULQVIRGILHNGASVLTQIGISWLGIRSGELGALIHNTILC 461
Db 408 PNMMDVFSVSNLVYIGRVLVSGSLTLIKQOIGISLFOFOSLKEISAGNIYITDNSNIC 467
Qy 462 FHTYPMDOLEFNPQALHTANRPEDECEGGLACHOICARGCHGPGPTOCVNSQFL 541
Db 468 YHTITMTLFTSTINORIVIRNRKAKENCTABGMCNHLSSGCGPGPDQCLSCRPS 527
Qy 542 RQCEVECEVRQLGLPREVYNAHCLPCHRECP-ONGSVTCGPEADOCVACAHYKDP 600
Db 528 RGRICIESCNLYDEGEREPENGSIQECDCPCKEKMEDGLTCHGPEDMCTCKSHKDDP 587
Qy 601 PCVACPGSVKPDLSMPIMKPEDEGAQPCPCINTHSCVDLDKGC-----P 649
Db 568 NCVKCPDGLQGANSE--IFKYADPRECHPCHPNCTOCNGFTSHDCIYPTWGSHSTP 645
Qy 650 AQRASPLTSIVSAV--GILLVVLGVNFGILIKRQOKIRKTKMYRLOEITELVPLTP 708
Db 646 QHAR--TPL--IAAGVIGLFIYIVGLTAVVRRASIK--KRALKRFL--ETELVPLTP 700
Qy 709 SGAMPNOQMRILKETELKRVKVLGSGAFVYKGIWIPDGENVKIPVAIKVLENTSPK 768
Db 701 SGTARNQQLRIKETELKRVKVLGSGAFVYKGIWIPDGETVYKIPVAIKILNETTGRK 760
Qy 769 ANKELIDEVYVWAGSPVVSRLGICLTSTVQLVTQLMPYGCGLDHPENRGLSGDL 828
Db 761 ANVEFMDEALIMASNDHPLVRLGVCISPTIQLVQLMPHGLGLEVYEHKDNIGSQLL 820
Qy 829 LNMCIQIAKMSYLEDVRLVHRDLAARNVLYVSPNIVK1TDFGLARLIDIEFVYADGG 888
Db 821 LNMCIQIAKMSYLEDVRLVHRDLAARNVLYVSPNIVK1TDFGLARLIDIEFVYADGG 880
Qy 889 KVPIMKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 948
Db 881 KVPIMKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 940
Qy 949 QPPTCTIVYMMVCMWIDSECRFRFRLVSEFSRMANDPORFVYIQNED--LGPASPLD 1007
Db 941 QPPTCTIVYMMVCMWIDSECRFRFRLVSEFSRMANDPORFVYIQNED--LGPASPLD 1000
Qy 1008 STFYRSLLEDMDGLVDAEEYLVPOQGFCDPAPAGAGMWHHRSSSTSGGDLTL 1067
Db 1001 SKFPNLDEEDLEMDMAEELVLP-QAFNITPP-----ITSPRRLDSNNS-----EI 1048
Qy 1068 GLEPSEEDAPRS-----PLAP--SEGAGSDVPFGDLGM 1098
Db 1049 GHSPPPAYTPMSGNQFYVRDGGFAAEGVSVYRAPATSTIPEAPVAGATATIFPDDSCN 1108
Qy 1099 GAKGLQSLPTHDPSPLOKYSDDPTVPLPS-----ETDGYVAPLTCSPQPEVYNQPDV 1151
Db 1109 GTLRPVAPVHVOEDSTQRYSDPTVFAPEBSRPGELDEEGYMTMDRDKPEQEYLVNPEV- 1167
Qy 1152 RPOPSRPEGPLPARPAGATLERAKTISPCKNGVGVKVFANCGAVENREYITPGGAAP 1211
Db 1168 -----ENPFVSRK-----KNGDLQ-----ALNDPEYNASNG--- 1194
Qy 1212 QPHPPA-----FSPAFDNLVYWDQDPERPG 1237
Db 1195 ---PPKADEYVNEFLYNTFANTLGAKEYLKNLITSMPEKAKKAFQNDPQVNHSLPERS 1251
Qy 1238 A--PPSTFKGTPT-----AENPEYL 1255
Db 1252 TLQHPDYLVQEYSTKYFYKNGRIRPIVAENPEYL 1285

```

```

RESULT 7
ERR4 RAT STANDARD: PRT: 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Babiaga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Fromme P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
RN [4]
RP FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN. NRG-
RP 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, RETICULIN AND
RP NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
RP NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AF041838; AAD08899.1; -.
DR EMBL, U52531; AAC53051.1; -.
DR HSSP, P11362; IFCG.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Ratin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP_2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_Pkinase; 1.

```


Db 1195 ----PPKADDEVNEPDLNTFTNALGNAEYMKNSILSVPEKAKKAFDNDPDYWNHSLPPR 1250
Qy 1237 GA--PPSTFKCTPT-----AENPEYL 1255
Db 1251 STLGHDPYLQGYSTKFKYKQNGRIRPIVAENPEYL 1285

RESULT 8
XMRK_XIPMA STANDARD; PRT; 1167 AA.
AC P1338;
DT 01-JAN-1990 (rel. 13. Created)
DT 16-OCT-2001 (rel. 40. Last sequence update)
DT 15-JUN-2002 (rel. 41. Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TV.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Witbrodt J., Adam D., Malitschek B., Mauelel W., Raulf F.,
RA Telling A., Robertson S.M., Schartl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
inducing Tu locus in Xiphophorus.",
RL Nature 341:415-421 (1989).
RN REVISION TO 515.
RL Schartl M.;
RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X16891; CAA34770.2; -
DR PIR; S06142; S06142.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.
FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 666 1167 POTENTIAL.
FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 716 724 PROTEIN KINASE.
FT BINDING 743 743 ATP (BY SIMILARITY).
FT ACT_SITE 835 835 ATP (BY SIMILARITY).
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793B0749DC1D55A CRC64;

Query Match 39.4%; Score 2695.5; DB 1; Length 1167;
Best Local Similarity 45.1%; Pred. No. 3e-137;
Matches 576; Conservative 163; Mismatches 387; Indels 151; Gaps 28;

Qy 4 AALCGWGLLLALPPGAAS-----OVCGTGDMKRLPASPETHLMDLRLYGGCGVQGN 59
Db 8 AALLD--LLLVISTSRCCSTDPDRKVCQSTSNQMTM---LDNHVLMKMKMYSGCNVVLEN 62
Qy 60 LELTYLPNTASLSFLQDIOEVGYVLIANQVRQVPLQRLRIVRGTQLFEDNYALAVLDN 119
Db 63 LEITYGENDLSFLQSIQEVGYVLIAMNEVSTIPLVNKLRIKRONLYEGNFTLLVMSN 122
Qy 120 --GDPINPNFTVSFWLRVPRKVSASHLEQLR--SLTEILKGVLIQRNPQLCYODITLM 175
Db 123 YQKNPSS-----PDVYGVGLKQLQLSNLTILSGGVKVSNNPLLCNVEITIMW 169
Qy 176 KDIFPHNNQALTLIDTRRSRAHCHPCSPMKGSRWCGSSSECCGLTITVCAGGC--ARCK 234
Db 170 WDIYDKTISNPTMNLPHAFERQCKQCDHGCYVNGSCWAPGHCQKFTKLLCAEQGNRCR 229
Qy 235 GPLPTDCCHEOCAAGCTGPKHSDCLAFHNSGICELHCPALVYNTIDTFESMNPGR 294
Db 230 GPKPIDCNEHCAGGCTGPRATDCLACDFDDGCKQTCPPKITYDIVSHQVNNPIK 289
Qy 295 YTFGASCTTACPVNTYLTSDVSGCTLVCPHINQEVTAEDGTORCEKSRPCARVCYGLME 354
Db 290 YTFGACVKECPSPNYVYVE--GACVRSASAGMLEVD--ENGKRSCKRCDCVCPVCGIGIG 347
Qy 355 HLREVRATVSANIOEPAGCKKIFGSLAFSPSGCDPASNTAPLPDPEQLVFEITLEITG 414
Db 348 SLSTTIANSNTNIFSISCTKINDIILNRSFSDPDHYKIGTWDPEHLMNLTYVKEITG 407
Qy 415 YLYISAMPDSLPLDLSVFQNLQVIRGRIILHNGAVS--LTQGLGISWLGIRLSREIGSLAL 473

```

Db      408 YLIVIMWPNMTSLSVFQNLRIIRGRTFSRGFSFVVVQVRHLQGLSLKSEVSAAGNY 467
Qy      474 IHHNTHLCVHTVPMQQLFRNPHQALHTANRDEDECVBGLACHOLCARGHWGSGPPQ 533
Db      468 LKNTLQLRANTIMRRLFRSEDOSELEYART-----ENQTCNNECSDDGCGGPTM 520
Qy      534 CYNCSQFLGQECVECECVLYLQGLPREVVAARHCLPCHPECOPONGSVTCFGEPAQCVAC 593
Db      521 CVSCHLVDRGRCVASCNLLQGEPRBAQVDRGVQCHQDEGLVDTDLCTCGPRAPACSKS 560
Qy      594 AHYKDPFCVARCSPGVKBDLSYMPIWKPREDGACQPCPINCTHSCVDLDDKCPAEQR 653
Db      581 AHFODRPOCIPRCPHGILGSDTL-IMKYADKMGQCORCHQNTQCSGSGGLSGCRGD-1 638
Qy      654 ASPLTIVSAVVCILLVVLGVVFGILIKRQOKIKYTMRLLOSTELVEPTTPSGAMP 713
Db      639 VSHSILAVGLVSGLLTIVALLIVLLRRRRRIK-KRRTIRCLLOEELVEPTLPSQAP 697
Qy      714 NOAMRILKETELRKVLSGAFGTVYKGIWIPDENVKIPVAIKVLRENTSPPKANKET 773
Db      698 NQAFRLILKETEKORVLSGAFGTVYKGLMNPDEGRIIPVAIKVLRATSPKXNOEV 757
Qy      774 LDEAVYVAGVSPYVRLIGICTSTVQVLTQMLPQGLLDHVENRGRIQSODLLNWCN 833
Db      758 LDEAVYVAVSDHPHVCRLIGICTLSAVQVLTQMLPQGLLDYVROHERICQWLLNMCV 817
Qy      834 QIAKGSYLEDVYLRDLAARVULVKSPTHVKITDQGLARLDIETENHAGGVYPIK 893
Db      818 QIAKGNVYLEERLVHRDLAARVULVKNPHVKITDQGLSKLLTADEKEYOAGGVYPIK 877
Qy      894 WMALESILRRFRTHQSDVSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPORPIC 953
Db      878 WMALESILQWYTHQSDVSYGVTVWELMTFGSKPIDGIRAKIASVLENGELRPPRIC 937
Qy      954 TIDVYIMVYKCMWIDSECRPRFELVSEFSRMAQDPQRFVYVIONEDLGAPSLDSTFYRS 1013
Db      938 TIEVYIMILKCMWIDSSRPREFELVSEFSQMARDSRYLVIOG---NLPSLDRRLFSR 994
Qy      1014 LLEDDMGDLVDAEELVLPQGGFFCPDRPACGAGVHHNRSSSTSGGDDLGLGLEPSE 1073
Db      995 LLSDD--DVVDDEYLLPYKRI-----NNGQS-----1020
Qy      1074 EEARSPRLPASEGAGSDVFDGDLGMGAKGLQSLPTHDPSPLOYSEDPV-PLPSETDG 1132
Db      1021 -----EPCIPPTH-----PRENSITLRNLSIDPQLNLEKDLG 1055
Qy      1133 YVAFLTCSPOPEYVNOQDVRRP-----PSPRE-----GLP-AARPGATLEDAKTL 1179
Db      1056 H-----EYVNOQGETSRSLSDIYNPNYEDLTDGWPVLSLQEAETVFSRPEYL 1105
Qy      1180 SGGKNGVVKVDFAFGAVENREYLLTPQCGAARPHRRPASFPDNLVYVDQDPRPERGAR 1239
Db      1106 NTNQNSL---PLVSSGMDPDY---QAG-----YQAAF-----LPQTAL 1140
Qy      1240 PSTFKCTPTAENPEYLG 1256
Db      1141 TGNQMFLLPAENLEYLG 1157

```

RESULT 9
 ERB3 HUMAN STANDARD; PRT; 1342 AA.
 AC P2160;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 15-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-precursor (BC 2.7.1.112)
 DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
 GN ERB3 OR HER3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Isling W., Miki T., Porgesu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Ploman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erb3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE p85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
```


KW	Alternative splicing.	
FT	SIGNAL	1 19
FT	CHAIN	20 1342
FT	DOMAIN	20 643
FT	TRANSMEM	644 664
FT	DOMAIN	665 1342
FT	DOMAIN	709 966
FT	NP BIND	715 723
FT	BINDING	742 742
FT	ACT SITE	834 834
FT	DISULFID	186 194
FT	DISULFID	190 202
FT	DISULFID	210 218
FT	DISULFID	226 235
FT	DISULFID	227 243
FT	DISULFID	231 243
FT	DISULFID	246 255
FT	DISULFID	259 266
FT	DISULFID	290 301
FT	DISULFID	305 320
FT	DISULFID	333 337
FT	DISULFID	500 509
FT	DISULFID	504 517
FT	DISULFID	520 529
FT	DISULFID	533 549
FT	DISULFID	552 565
FT	DISULFID	556 573
FT	DISULFID	576 585
FT	DISULFID	589 610
FT	DISULFID	613 621
FT	DISULFID	617 629
FT	CARBOHYD	126 126
FT	CARBOHYD	250 250
FT	CARBOHYD	333 353
FT	CARBOHYD	408 408
FT	CARBOHYD	414 414
FT	CARBOHYD	437 437
FT	CARBOHYD	469 469
FT	CARBOHYD	522 522
FT	CARBOHYD	566 566
FT	CARBOHYD	616 616
FT	VARSPLIC	141 183
FT		
FT	VARSPLIC	184 1342
FT	CONFLICT	560 560
FT	CONFLICT	1064 1064
SO	SEQUENCE	1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
Query Match 35.5%; Score 2427.5; DB 1; Length 1342;		
Best Local Similarity 40.5%; Pred. No. 8.1e-123;		
Matches 534; Conservative 192; Mismatches 453; Indels 141; Gaps 34;		
QY	10 GLLALLPFGAA--STVCTGTDMLRLPASEPHILDMRLHYOGCQVQGNLELTYLPT	67
DB	11 GLTSLARGSEVGNQAVCPGTNLGISTVGAENQYQTLXKLYERCEVVMGNLEIVLTS	70
QY	68 NASISFLDIOEVQGVLIHNOVROPLORLRIYRGOLPEFDNVALDLDGDDLANNN	127
DB	71 NADISFLDMIREVTIYLVANNEFSTLPLNLRVVRGVOVYDGKRAIFV-----MLNTN	124
QY	128 NFTVSWLVRPKVSNH--LEOLR--SLTEILKGVLIGRNPOLCYQDTITLMKDIPIKINO	184
DB	125 -----TNSSHALRQLRLTQLTEIILSGVYIEKNDLCHMDITIDMRDIVRDRD-	171
QY	185 LATLITIDNRSLACHPGSPMKSGKNGESSDQSLRTTVACAGC-ARCKAPLPTDCH	243
DB	172 --AEIVLVNDNGSCPCPCHEVCKG--RCWGPGESEDCTLTCTICAPONGHCFCFNPNOCH	228
QY	244 ECGAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDPFESNPNEGRTFGASCYT	303
DB	229 DECAAGCGGPODITDFACRHFNDSGACVPRCPQPLVYNKLFIPLEPNPTTKYOGGVCA	288

QY	304 ACPYNYLSTDVSGCTLVCEPLHNOETABDGTORCEKSKRCARVCYGLGMEHLREVAVT	363
DB	289 SCPHNFV--VDQTSVCRACPRPDMKEDV--KNGLMMCEPCGGSLCPACAGTGG--SRGTVD	344
QY	364 SANIOEFACCKKIFGSLAPLPSFGDDPSASNTAPLOPOLQOFLETLEITGYLYISAMP	423
DB	345 SSNIGFVNCITLGLDPLITGLNDGPMHKIPALPEKLNFRVREITGYLINISWPP	404
QY	424 SLPLDSVPONLQVINGRILHNGAYS--LTLOGISWLSGRSRELQSLALIHNNHLCF	482
DB	405 HHNHSVFENLTTTIGRSLYNGFSLINKNLNVSLGRSLKEISAGRIYISANPOLCY	464
QY	483 VHTVPMDOLEFRPHOALLATA--NRPEDECVGEGLACHOLCARHNGCPPTQVNCNOLF	541
DB	465 HHSIMMTKYLRPTEBERLDIKNNRPRRDCVAAGKVCDPICSSGCGCPGPGCGLSRNVS	524
QY	542 RQECVECEGRVLQGLPREYVNAHRLCPHPECOPONGSVTCGPEADQCVAAHYDPPF	601
DB	525 RGVGVCTHCPNFIENGEPFAHBAECFSCHECOPMEGTATCNGSGSDTCAQCAMFRDGRH	584
QY	602 CVARCPGKPPLSYMPIKFPDEBGAOCPICNTHSCVDLDDKCCRAEGR-----SPL	657
DB	585 CVSSCPHVLG--AKGPYKYPDVQNECPCHENCTQCGKPELODCLQTLVLIGKTHL	642
QY	658 TSIVSAVVGILLVVLGVFGILIKRQOKIR--KYTMRLQSTELVEPLTPSGAMPNOA	716
DB	643 TVALYVING--LVTFMMLIGTFVYWRGRIONKAMRKYLERGSEITEPLDPS--EKANKV	699
QY	717 QMRILKETELRVKVLGSGAFSTYKGIWIPGENYKIPVAIKVLENTSPKANKELDE	776
DB	700 IARIFKETELRLKVLGSGVFTVHKGVWIPGESIKIPVCIKVIDKSGRQFQAVTDH	759
QY	777 AVYMAQVSPYVSLGICLTSTVQLYQLMAYGCLLDHYRENRGLSGODLLMCMQIA	836
DB	760 MLATSLDHAIVRLGLCPGSSLOLVQYLPGLSLDVRHGRGLAQGLLLMNVQOIA	819
QY	837 KCMSTLEDVRLVHRDLAARNVLYKSPNHVYKIDTFGLARLLDIDETEVHADGKVPYIKMA	896
DB	820 KCMYTLBEHGMVHRDLAARNVLYKSPQVQVADLPDDKQLLYSEAKTPYIKMA	879
QY	897 LESILRRFTHOSDVWSGVTVWELMTFGAKPYDGI PARIEPLDEKGERLPQPICTID	956
DB	880 LESIHFGKTYTHOSDVWSGVTVWELMTFGAEPYAGIRLAEVPLDLEKGERLQAPQICTID	939
QY	957 VYNIWVKCMIMISECRPRRELVSFSPARAPQPFVYIQNDLGRA---SPLDSFTFYS	1013
DB	940 VYMWVWKCMIMIDENIRPTFEKLANEFTRMARDPPRYLVIKRES--GGGIAAPGEPRGLTYNK	998
QY	1014 LLEDDMDGLVDAEYLVPOQGFCEPDRPARGAGVWVHNHRSSSTRSGGDLTIGLEP--S	1072
DB	999 KLEVELEPELDLDLDLEAED-----NLATTTLSALSILPVGTIN	1039
QY	1073 EEEARSPPLAPSEGASDVFDGLGMGAQGLQSLPTHD--PSPLORYSEDPVPLP----	1127
DB	1040 RPRGQSILSPSSGY--MPNQGNLGESSQESAVSSSSEKCRPRVSLH-----PMRGL	1092
QY	1128 --SETDGYVA-----PLTSGQPF-----YNOQPDVPRPPSPBSCP----	1162
DB	1093 ASESSEGHVTSGEALQEKVSMCRSRSSRSPRGDSAYHQRSLTLPPVPLSPGLE	1152
QY	1163 -----LPAARPAATLERAKTLSF--GKNQV-----KQVFAFGAVENPEYLTPOG	1207
DB	1153 EEDVNVGYWMPDHLKGTSSREGTSSVGLSSVLTGEEDD-----EEYVNNRR	1204
QY	1208 GAAPRPPAPAFSPAFDNLVYWD-----QDPRERGAPOSTFKGPTAENPEYL	1255
DB	1205 RHSP--PHPPRPSLELEGEYMDVDSLSASLGTSQSCPLHVPPLMPRTAGTTPDDOYEYM	1263

RESULT 10
 EBB3_RAT
 ID_EBB3_RAT
 AC_Q62759; Q62955;
 STANDARD; PRT; 1339 AA.

DT	15-DEC-1998 (Rel. 37, Created)
DT	15-JUN-2002 (Rel. 41, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE	(c-erbB3).
GN	ERBB3.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Sprague-Dawley; TISSUE=Liver;
RX	MEDLINE=96096535; PubMed=8522190;
RA	Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT	"Cloning of the rat ErbB3 CDNA and characterization of the
RT	recombinant protein.";
RL	Gene 165:279-284(1995).
RN	(2)
RP	REVISIONS TO 85; 513 AND 565.
RA	Hellyer N.J., Koland J.G.;
RL	Submitted (DEC-2001) to the EMBL/Genbank/DBSJ databases.
RN	[3]
RP	SEQUENCE OF 922-1097 FROM N.A.
RC	STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX	MEDLINE=97184212; PubMed=9030624;
RA	Carroll S.L., Miller M.L., Frohnet P.W., Kim S.S., Corbett J.A.;
RT	"Expression of neurogulin and their putative receptors, ErbB2 and
RT	ErbB3, is induced during Wallerian degeneration.";
RL	J. Neurosci. 17:1642-1659(1997).
CC	-1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC	tyrosine phosphate.
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC	(POTENTIAL).
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC	SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC	-1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC	AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC	PHOSPHATIDYLINOSITOL 3-KINASE.
CC	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U29339; AAC28498.2; -.
DR	EMBL; U52530; AAC53050.1; -.
DR	HSSP; P11362; IFGK.
DR	InterPro; IPR000494; EGFR_L domain.
DR	InterPro; IPR000719; Euk_Pkinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_Pkinase.
DR	Pfam; PF00069; Pkinase; 1.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF01030; Recep_L domain; 2.
DR	PRINTS; PR00109; TYRKINASE.
DR	ProDom; PD000001; Euk_Pkinase; 1.
DR	SMART; SM00261; FU_5.
DR	SMART; SM00219; TYRK; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT	SIGNAL 1 19
FT	POTENTIAL.
FT	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT	DOMAIN 20 643
FT	EXTRACELLULAR (POTENTIAL).

Query	Match	Similarity	Score	DB	Length
Db	3 LALCRGGLLTLALPPGAA---	34.4%	2357.5	DB 1	1339
Qy	60 LELTYLPTNLSLFLQDIQEVGYVLI	40.7%	2357.5	DB 1	1339
Db	7 LQVLC-----FLSLARGSEKMSQAV	40.7%	2357.5	DB 1	1339
Qy	63 LEIYLGTGNADLSFLQWIREVTGV	40.7%	2357.5	DB 1	1339
Db	120 GDLPTNNENFTVSMFLNPKVSASH	40.7%	2357.5	DB 1	1339
Qy	120 ---MLNTN-----TNSSHALRQL	40.7%	2357.5	DB 1	1339
Db	177 DIFKNNQALTLIDTNRSPACHSC	40.7%	2357.5	DB 1	1339
Qy	165 DIVAVR---GAEIYVKNKNGANCP	40.7%	2357.5	DB 1	1339
Db	236 PLPLPDCCHQCCAGCTGPKASDCL	40.7%	2357.5	DB 1	1339
Qy	221 PNPQOCHDEACGSGSPQDDTCAC	40.7%	2357.5	DB 1	1339
Db	296 TFGASCTYACRYNLTSDVGSCT	40.7%	2357.5	DB 1	1339
Qy	281 QYGVGVCAVACPHNV-VDQTFCA	40.7%	2357.5	DB 1	1339
Db	356 LREYRAVTSANIOEFACCKKIFG	40.7%	2357.5	DB 1	1339
Qy	338 -SRQYVDSSNIDGFVNCTKILGN	40.7%	2357.5	DB 1	1339
Db	416 LYSISAWPDSLPDISVFPNLQVTR	40.7%	2357.5	DB 1	1339

```

Db 397 LNIQSPHMHNFVSFNSLTTIGSLNRGSLIMKNLNTSLGFSLEISGRVYI 456
475 HHNTHLCVHTVPMDFRNPHQALLHTA-NRPEDEVCEGLACIQLARCGHCHWPGPTQ 533
457 SANQOLCVHSHLNTWRLRGRPSBERLDIKYDRPLCECLAEGRVCDPLSSGGCWPGRQ 516
534 CVNCSQFLRGCECEECRVLOGLPREVYNARHCLPCHPECORPONGSVTCFGEADQVAC 593
517 CLSCRNVSRGVCTHNCNFCLOGEPRFVHEAQCFSCHEPCLMEGTSTCNCSGSDACARC 576
594 AHYDPRPCVARGSGVPRDLSTYMPIMKPEDEEGACQCPINCHTSC--VLDLDDGCAE 651
577 AHFRDGRPCVNSCPHGLG--AKGTIYKPPDAQNECRCHENCTGCGNPELDLQAE 634
652 QRASPLTIVSAVVGILLVVLGVVFGILIKRROKIR-KYTMRLLOETLEVEPLTPSG 710
635 VLKSPHVLIVAVTG--LAVILMLIGSFLVWRGRIONKAMRYLBERGSEIEPLDS- 691
711 AMPNQAOMRILKETELRKVYLVGSGAFSTYKGIWIPGENVKIPVAIKVIRENTSPKAN 770
692 EKANKVLAIRIFKETELRKVLGSGVFGTVHKGIMIPEGESIKIPVCIKVIEDKSGRSF 751
771 KEILDEAVMAGVSGSPVSRLLGICLTSTVOLVQLMRYGCLLDHVRNRLSGDILN 830
752 QAVIDHMLAVGSLDHAHIVRLGLCPGSSLDLVQYLLGSLDHYKHRETLGPOLLN 811
831 WCMQIAGKMSYLEDVRLVHRDLAARVYLVKSPNHVKITDFGLARLLDIDETEHYADGKV 890
812 WGVQIAGKMYLLEHSHVHRDLARNWLKSPQOVADLPPDDKQLHSEAKT 871
891 PIKMMALLESILRRPFTHOSDVSYGVTVWELMTFGAKPYDGI PAEIPDLLEKERRLP 950
872 PIKMMALESIHFGKYTHOSDVSYGVTVWELMTFGAEYAGIRLAIEIPDLLEKERRLP 931
951 PICTIDVYVMVKCMIDSECRPRELVSEFSRARPORQFVYIQNDLDPASLDSFT 1010
932 QICTIDVYVMVKCMIDENIRPTFKELANETRAARPPRYLVYKRAS--PGRTP--PAA 988
1011 YRSILLEDDMDGLVDAAEYLVPOQGFPCPDAPAGAGVHHRHSSSTRSGGDLTGLE 1070
989 EPSVLTKTEL-----QEAELEPEL-----DLDLLE 1014
1071 PSEF-----EAPRSPAPSEG-----AGSDVPDGLDMGAK 1102
1015 AEEEGLATSLGALSLLPTGTLTRPRGSQLSPSSGYPMKMOSSIGEAACLAVLGSREQ 1074
1103 GLOSLPHDPSPLRYSRDPVPLPSEIDGV-----APL-----TC-----SPOPE-- 1144
1075 FSRPISLH-PIPRGR-----PASESEGHVTGSEAELOKVSVCRSRSRSPRPRGD 1126
1145 --YVNPDPVRPQPSPREGP-----LPAARPAATLTERAKTLSP-GKNGVY--- 1167
1127 SAHNSQSHSLITPTPLSPLEEDDNGYMPDTHLGRASSRBEGLTSSVGLSLVLTGE 1186
1188 ---KDVFAFGAVENPEYLPFGGAAPQHP 1216
1187 EEDDED-----EEYEYMKRKRGRSP-PRPP 1209

```

RESULT 11

```

ID EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (EGfr)
DE (Gurken receptor) (Tropo protein) (Drosophila relative of ERBB).
OS EGFR OR TOP OR C-ERBB OR DER OR CG10079.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chame M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abilil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baev A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Butts K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA LaRoche J., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matcei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nuskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 (7)
 RN SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=Daekanyeon;
 RX MEDLINE=85137938; PubMed=2983232;
 RA Wadsworth S.C., Vincent W.S. III, Bildeau-Wentworth D.,
 RT "A *Drosophila* genomic sequence with homology to human epidermal
 RT growth factor receptor,"
 RL Nature 314:178-180(1985).
 (8)
 RN SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RA MEDLINE=92038942; PubMed=1936959;
 RX Raz E., Schejter E.D., Shilo B.Z.,
 RT "Interallelic complementation among DER/rlb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases,"
 RL Genetics 129:191-201(1991).
 (9)
 RN REVIEW.
 RP MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.,
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
 RT EGF receptor,"
 RL Cell 89:13-16(1997).
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES. AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UNBOUTHOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANTENNA OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; -;
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAAS1462.1; -;
 DR EMBL; K03417; AAAS1460.1; -;
 DR EMBL; K03416; AAAS0965.1; -;
 DR EMBL; K03418; AAAS1461.1; -;
 DR EMBL; AF109077; AAD26134.1; -;
 DR EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -;
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -;
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; -;
 DR EMBL; AE003454; AAP46732.1; -;
 DR EMBL; X02293; CAA26157.1; -;
 DR EMBL; X78920; CAA55523.1; -;
 DR EMBL; X78918; CAA55521.1; -;
 DR EMBL; X78919; CAA55522.1; -;
 DR PIR; A00640; GOFPE.
 DR HSSP; P11362; TECK.
 DR Flybase; FBgn003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recp_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR Prodom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; TYKIC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.9%; Score 1977; DB 1; Length 1426;
 Best Local Similarity 32.9%; Pred. No. 1.2e-98;
 Matches 470; Conservative 185; Mismatches 431; Indels 344; Gaps 42;
 Oy 24 QVCTGTDMKRLPASPETHLDMLRHLVYGGQVGVGNLELYLPT-NASLSFLODIOVOG 82
 Db 100 KICIGTKSRLLVPSNKEHHYRNLDRTYVNDGKLTWDPNENLDSFLDNIREVTG 159
 Oy 83 YVLAHNVQVRPLQRLRIVRGTOF-----EDNYALAVLDNGPLNNFNFTVSFWLRV 137
 Db 160 YILISHVDVKKVPKQIIRGRTLFSLVSEEEKAL-----FVTVSKMYTLEI 208
 Oy 138 PKVSASHLEQLRSLEILKGVLLQRPQLCYQDTILMKDIFKRNQALALLIDNRSRA 197
 Db 209 P-----DLRDVNLQGVGFHNNYNLQMKMTIQMSEIVNSGTDVYVYDFAPBRE 257
 Oy 198 CHPCSPMKGRCGCGSESDCOSLTRVCAGACA--RKGPLPDQCHGEOCAAGTGPKH 255
 Db 258 CPKHESCTHG-CHGEGPKNCKQSKLCSQACAGRGYGRKPRECHLPACAGCTGPTQ 316
 Oy 256 SDCLACLHFHNSGICELHCPALVTYNTDFESMPNPEGRYTFGASCYVACENYVLTSDVG 315
 Db 317 KDCIACKNFFDEAVSKSECPMRKYNPTTYVLETNPEKGAVYAGATCVKECP-GHLIRNDG 375
 Oy 316 SCLIVCPPLHNOEVAEDQTCRCGSCRCPCARVYCGLGMEHLREVRATVSANIQFAGCKK 375
 Db 376 ACVRSCPODDKMDKGE-----CVPCKGCPPTCGGVYVLA-----AGNIDFRNCTV 422
 Oy 376 IFGLAFLPESFDG--DPASNTA-----PLQRPQLQVFETLEITGYLYISAPDLSLD 427

```

423 IDGNRIIDQTSFGQDVYANTWGPRIYPLDPERREVSITYKETGTGYNIEGTHPQERN
428 LSVFONLQVIRIRILANGAY-SLTQGLIGISWLGRLSRLSGSLALIHNTLGFVHTV
483 LSYFNTLETIRQIMESMFALALVKSLVSLERNLKQISSGVVIOHNDLCLVSNL
487 FMDQLFRNPQALHTANRPEDECVEGLAQHQLCARQGWGPPTQCVNCSQPLRGDEC
543 RMPALOKEPEOKVWVNNENLRDLCEKNGTICSDQCNEDGCMGAGTDQCLTCNFPNFGTC
547 VEECGVLQGLPREYVNAHCLPCHPECCQONGSVTCFGEAOQCVACAHYKDPFCVARG
603 IADCGYISNAYK--FDNRTCKICHEPCR-----TCNGAGADHCOECVAVRDGHCVCSEC
607 P-----SGVK-----PDL 614
655 PKNKXNDRGVCRECHATCGCPKPDITIGACTTGNLAINNDATVAKCLAKDKCPD- 713
615 SYMPLWKF--PDEBACOP-----CPI-----NCTH----- 638
714 GY--FMEEYVHQEQSLKPLAGRAVCRKCHPLCELCITVYHGEQVCSKCTHYKREQCEET
639 -----SC-----VDLDDKG----- 647
772 ECPADHYDEORECFQHRPECNCTGCGEADDCSKCRNFKLFDANETGYVNSTMFNCTS 831
648 -CPAEOR-----ASPLTS-----IYSAVAGILVVLGVFGI 679
832 KCPLEMRHVNYQYTAIGPYCAASPSSKITANLDVNMIFITGAVLVETIILCV--T 889
680 LKRRQOKIRKYT--MRRLQETELVEPLTSGAMPNOQMRLKETELRKVKVLGSGAF 737
890 YICROKQAKKETVMTALSGCEDSEPLRPSNIGANCKLRIVDAELRKGVLMGMAF 949
738 GTVYKGIPIPGENKIVAIKVLENTSPKANKIILDEAVYMAVGSPYSRLIGICLT 797
950 GRVYKGVAVPSEENKIVPAIKELKSTGAESESEFLEAYIMAAEEHVNLIKLAAYCMS 1009
798 STVOLVTOIMPYGCLLDHVRNENRGLSGODLIMCMQIAKGSYLEDVRLVHRDLAANV 857
1010 SOMMLITQMLPGCLLDVVRNNRDKIGSKALLNWSQTOIAKGSYLEEKRLVHRDLAANV 1069
858 LVKSPNHVYKIDFGIARLLIDETVYHADGKVPKIMMALESILRRRPTHOSDVMSYGT 917
1070 LVQTPSLVKITDFGLAKLLSSDSNEYKAAAGKMPKIMLALBECIRNRVFTSKSDVMAFGVT 1129
918 VMEIMTFCAGKPYDGIAREIPDLLEKGERLPQPCITIDVYIMKCMIMDSECPRRERE 977
1130 IMELTFQORPHENIPAKDIPDLLEVGLKEQPEICSDIDICTLLSCWHLDAAMPRTFKQ 1189
978 LVSESRMARDDPQRFVVIQNEDLG--PASPLDSTFYRSLLEBD--DMGDLVDABEYLV 1032
1190 LTTYPAFAFARDPGRRLAIPGDKFTRLPA-----YTQDEQDLIRKLAFTTDSSEALAK 1242
1033 QCGFFCPDPAPQAGCMVHRRSSSTRSGGDLTLGLEPSEBEAD-----RSLPAPSEG 1086
1243 PDDYLOPFAAPGSPS-----HRTDCT-----DEMFKLNRKYCDPSPKNSS 1281
1087 AGSDVFDG--DLGMAKAGLSLPTHPSPLQKRSDEPTYPLBSETGYAPALTCSPQ 1143
1282 TGDDERDSSAREVGGNLR-----LDLPDEDDYDIMP--TCQGGP 1319
1144 EYVNPQDVRRPOPSPBREGPLPAARPAAGATLBRAKTISGKXGVVDVPAFGAAVNPXYL 1203
1320 NNNNNNN-----NPNQNNMAAVGAAGYM-----DLGCVPSVDNPEYL 1358
1204 ----TPOGGAAPQH-----PPAPSP-AFDNLYYWD 1230
1359 LNAQTLGVSPEPIPTOTIGIPVMGPGTMEVKVPMGPSEPTSSDHEYYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (rel. 01, Created)
DT 15-JUN-1999 (rel. 38, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN v-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_Taxid=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor".
RL Cell 41:719-726(1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INDUCED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A06643; TVCHLV.
DR PIR; B06643; TVFVLV.
DR HSSP; P1362; IFGK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR Prodom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KM Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146
FT BINDING 165 165
FT ACT SITE 257 257
FT SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 7,6e-87;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
OY 593 CAHYKDPFCVACRPSGVKPLSYMPKPFDEBACQPCPINTGHSQVDDLDKQCPAQ 652
DB 3 CAHFIDGPHCVKACPAVAGENDTL-VWKYADANAACVQLCHPNCRTRGCKGPLEGCP-- 58
OY 653 RASPLTGISVAVV-GILLVVLGVVFGILIRROOKIRKYMRRLQETELVEPLTSPGA 711
DB 59 NGSKTPSIAGVAVGGLCLLVVAGIGLYLRR-HYVAKRTLRLDLERELVEPLTSPGE 117
OY 712 MPNOQMRLKETELRKVKVLGSGAFVYKGIPIPGENKIVPAIKVLENTSPKANK 771

```

```

DB 118 APOHRLRLKETEFKVKVVLGSGAGFYKGLMIPEGEVKIPVAIKELREATSPKANK 177
OY 772 ELIDEAYVWAGSPVYSRLGICLTSTYQVLTQMLPYGCLLDHVENGRGLGSDLLMW 831
DB 178 ELIDEAYVWASVDNPNVCHLGLCTSTYQVLTQMLPYGCLLDHVENGRGLGSDLLMW 237
OY 832 CMQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLDDIDETEHYADGKYP 891
DB 238 CQIAKGMVLYEERLVRHDLAARNVLYVTPGHVKITDGLAKLGLADEKEHYAEGKYP 297
OY 892 IKMALESLIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
DB 298 IKMALESLIRHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSYLEKGERLP 357
OY 952 ICTIDVYIMVCMWIDSECRPFRELVSFESRMADPQRFVYIQ-NEDLGASPLDSTF 1010
DB 358 ICTIDVYIMVCMWIDADSRPFRELVSFESRMADPQRFVYIQDERHMLPSPDSKF 417
OY 1011 YSLLEDDMDGLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLE 1070
DB 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
OY 1071 PSEERAPSPFL-----APSEGASDVFDGDLGMAKGLQSLPTHPSPLOKRYSEDPTVP 1125
DB 450 -----SRTPLLSLSATSNMSTNCID-----RNGGSHVREDSPFORYSPTGN 495
OY 1126 LPSET--DGTVAPLTCSPQPEYVNOQDVRPOPSPREGPLPARPAGATLERAKTUSPOK 1183
DB 496 FLEESIDDFL-----PAPEYVNO--LMPKRS-----TAMQ 526
OY 1184 NGVAKOVF-----AFGGAIVENPEYLTPOGGAQDPHPPAFSPAFDNLVY 1228
DB 527 NQIVNNISLTALSKLPMDSRYQNSHSTAVDNPEYL-----NTNQSPILAKTVFESSPY 578
OY 1228 WQO-----DPEF-----RGAPSPFKGPTPTENPEYLGDLVP 1260
DB 579 WIQSGNHQINLNDPDYQDPLNETRKNGLLKVAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=84223957; PubMed=6328658;
RA Debnire B., Henry C., Benaisa M., Biserie G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.

```

```

CC -I- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; 1FCK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW transferase; tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146
FT BINDING 165 165
FT ACT_SITE 257 257
FT CONFLICT 29 29
FT CONFLICT 140 140
FT CONFLICT 146 146
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D009 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 2, 2e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

OY 593 CAHYKDPKRCVAPCSGVKPDLSYMPIMKFPPEBEGCCQPCPINCHTSCYVDLDDKCPAQ 652
DB 3 CAHFIDGPHCVACPCPGVIGENDTL-VKRYADANAVCQLCHNPTCRGCGPGLGCP--- 58
OY 653 RASPLTISYSAV-GILVAVGVVFGILIKRROKIRYTRRLRLOETELVEPLTPSGA 711
DB 59 NSKTPSIAGAVVGGCLLVVVGIGLVLRRR-HYRKRTLRLLQDELEVEPLTPSGE 117
OY 712 MENOQMRILKETELAKVVLGSGAGFYKGLMIPEGEVKIPVAIKELRENTSPKANK 771
DB 118 APOHRLRLKETEFKVKVVLGSGAGFYKGLMIPEGEVKIPVAIKELRENTSPKANK 177
OY 772 ELIDEAYVWAGSPVYSRLGICLTSTYQVLTQMLPYGCLLDHVENGRGLGSDLLMW 831
DB 178 ELIDEAYVWASVDNPNVCHLGLCTSTYQVLTQMLPYGCLLDHVENGRGLGSDLLMW 237
OY 832 CMQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLDDIDETEHYADGKYP 891
DB 238 CQIAKGMVLYEERLVRHDLAARNVLYVTPGHVKITDGLAKLGLADEKEHYAEGKYP 297
OY 892 IKMALESLIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
DB 298 IKMALESLIRHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSYLEKGERLP 357
OY 952 ICTIDVYIMVCMWIDSECRPFRELVSFESRMADPQRFVYIQ-NEDLGASPLDSTF 1010
DB 358 ICTIDVYIMVCMWIDADSRPFRELVSFESRMADPQRFVYIQDERHMLPSPDSKF 417
OY 1011 YSLLEDDMDGLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLE 1070
DB 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
OY 1071 PSEERAPSPFL-----APSEGASDVFDGDLGMAKGLQSLPTHPSPLOKRYSEDPTVP 1125

```

```

Db      450 -----SRPILSLSLSSATSNNSATCID-----RNGQGHPRVDESFVQRYSSDPDIGN 495
Qy      1126 LPSET--DGYVAPLTCSPQPEVYVNPDPVPPSPREGEPLPARPAGATLEBAKTLSPGK 1183
Db      496 FLEESIDGFL-----PAPEVYNQ--LMPKPSSTAM-----524
Qy      1184 NGVAVDVFVF-----GCVENPEYLTQGGAAAPQPHPEPASPAPD 1224
Db      525 --VNGQIINFIISLTALSKLPMDSRQNSHSTAVNDPEYL-----NTNOSPILAKTVFE 574
Qy      1225 NLYYWDQDPPERGAPSTFKGTPTAENPEY 1254
Db      575 SSPYWIQSGNHQ-----INDNPDY 594

RESULT 14
ERBB_AVIEU
AC      P11273;
DT      01-JUL-1989 (Rel. 11, Created)
DT      01-JUL-1989 (Rel. 11, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
V-ERBB.
OS      Avian erythroblastosis virus (strain ts167).
OC      Viruses; Retrocid viruses; Retroviridae; Avian type C retroviruses.
OX      NCBI_TaxID=103898;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=87064458; PubMed=2878364;
RX      Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT      "A single amino acid substitution in v-erbB confers a thermolabile
RT      phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT      cells.";
RL      Mol. Cell. Biol. 6:1751-1759(1986).

-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
-!- tyrosine phosphate.
-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M13179; AAA42401.1; -
DR      PIR; A25231; TVVEVB.
DR      HSSP; P11362; 1FGK.
DR      InterPro; IPR000719; Euk_kinase.
DR      InterPro; IPR001245; Tyr_kinase.
DR      Pfam; PF00069; pkinase; 1.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00219; TyrcK; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      TRANSFERASE; PS50011; PROTEIN_KINASE_DOM; 1.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132 399
FT      NP_BIND 138 146 ATP (BY SIMILARITY) .
FT      BINDING 165 165 ATP (BY SIMILARITY) .
FT      ACT_SITE 257 257 BY SIMILARITY.
FT      VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB) .
SQ      SEQUENCE 540 AA; 60412 MW; 5B53297AA06BB65D CRC64;

Query Match 23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Fred. No. 1.6e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
593 CAHYKDPFCVACRCSGVKPDLSYMPIKPFDEEGACQCPICNTHSGCVLDLDDGCPAQD 652

```

[illegible]

; Entered [jdelaval 22-Jul-03 6:11]
SEQ4-5-25-12
MELALICRMGLIALALPPGAASOVCTqYkanekfigitellYOGCOVVQNLLETYLPTNASLFLD
IOEVGVVLIANOVROVPLORLRIYRGTOLEFEDNYALAVLDNGDPLNNTTPTYGASPGGLRELOSLT
ELKGGVLIORNPOLCYODTILMKDI FHKNNQALTLIDNRSRACHPCSPMCKGSRGWSSESDOSIT
RTVCAGGACRCCKPLPTDCHEOCACCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTRESNPRE
GRYTGASCVTACPTNYLSTDVGSCTLVCPHNOEVTAEQTORECKSPCARVCYGLGMEHLREVRAY
TSANI OEFAGCKKIFGSLAFLPESFGDPSANTAPLOPEOLOVETLEETIGYLYISAMPDSL.PDLVSFO
NIQVIRGRLIHNAGAVSLTLOGIGI SWLGRSLRELSGLALIHNNHLCFVHTVPMDOLEFRNHQALHT
ANRPEDECVGEGLACHOLCARGCMGPGPTQCVNCSQPLRGOCVEECRYLQGLPREYVNAHCLPCHPE
COPONGSVTCFGEPEADQCVACAHYKDPFCVAPCSGVKPDLSYMPIMKRPDEGACOPCPINCTHSVD
LDKCPAORASPLTISIVSAVVGILLVVGLVFGILLIKRROKIRKYMRLLOETELVEPLTSGAM
PNOQMRILKETELERKVKVLGSGAFVYKGIWIPGENVKIPVAIKVLENTSPKANKELDEAVVMAG
VGSPLYVLSRLIGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGQDLNMCQIAKMSYLEDVRLVHRDL
AARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPIMKMALESILRRRFTHOSDWSYGVTWELM
TFGAPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVYKMWALDESGRPRELVSERSMARDPORF
VVIQNEEDLGASPLDSTFTRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGG
GDLTIGLEPSEEEAPRSPLAPSEGASDVFDGLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDQ
YVAPLTCSPQEPYVNOPODVAPQPPSPREGPLPARPAGATLERAKTILSPKNGVVKDVFAPFGAVENPEY
LTPQGAAPQHPHPPAFSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYIGLDVPI

; Entered [jdelaval 22-Jul-03 6:14]
SEQ4-5-25-14
MELALICRMGLIALALPPGAASOVCTfnfvcfwLrypkveashlelyOGCOVVQNLLETYLPTNAS
LSRIQIOVOCGVVLIANOVROVPLORLRIYRGTOLEFEDNYALAVLDNGDPLNNTTPTYGASPGGLRE
LOSLTEILKGGVLIORNPOLCYODTILMKDI FHKNNQALTLIDNRSRACHPCSPMCKGSRGWSSESD
DCOSLTRVCAGGACRCCKPLPTDCHEOCACCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFE
SMNPEGRYTFGASCVTACPTNYLSTDVGSCTLVCPHNOEVTAEQTORECKSPCARVCYGLGMEHL
REYVAVTSANI OEFAGCKKIFGSLAFLPESFGDPSANTAPLOPEOLOVETLEETIGYLYISAMPDSL.P
DLVSFONLQVIRGRLIHNAGAVSLTLOGIGI SWLGRSLRELSGLALIHNNHLCFVHTVPMDOLEFRNH
QALHTANRPEDECVGEGLACHOLCARGCMGPGPTQCVNCSQPLRGOCVEECRYLQGLPREYVNAHCL
LPCHPECOQONGSVTCFGEPEADQCVACAHYKDPFCVAPCSGVKPDLSYMPIMKRPDEGACOPCPINC
THSCVDLDKCGCPAEGRASPLTISIVSAVVGILLVVGLVFGILLIKRROKIRKYMRLLOETELVEPL
TPSGAMPNOQMRILKETELERKVKVLGSGAFVYKGIWIPGENVKIPVAIKVLENTSPKANKELDE
AYVMAGVSPYVLSRLIGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGQDLNMCQIAKMSYLEDVRL
LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPIMKMALESILRRRFTHOSDWSYGV
TWELMTPGAPYDGI PAEIPDLLEKGERLPQPICTIDVYIMVYKMWALDESGRPRELVSERSMARDPORF
RDBQRVVIQNEEDLGASPLDSTFTRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGG
STRSGGDLTIGLEPSEEEAPRSPLAPSEGASDVFDGLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDQ
YVAPLTCSPQEPYVNOPODVAPQPPSPREGPLPARPAGATLERAKTILSPKNGVVKDVFAPFGAVENPEY
LTPQGAAPQHPHPPAFSPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYIGLDVPI

1000
1000
1000

: Entered [jdelavel 22-Jul-03 6:21]
SB04-59-73-12
MELALICRMGLLIALLPFGAASOVCTGTDMKRLRPLASPETHLDMRLHLYOGCOVVOGNELETYLPFNAS
LSFLDIOEVOQVYkanskfigitieiQLRiVRGTOL.FEDNYALAVLDNGDPLNNTTPVTGASPGGLREL
QLRSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNOLATLIDTNRSRACHPCSPMKSGSRCS
DCOSLTRVCGGCGARCGPLPTDCCHOCAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTOTFE
SMNPGRYTFGASCVTACVYNYLSTVSGCTLYCPLHNOEVTADGTORCEKSGKPCARVCGYGLMBHL
REVRATYSANIOEPAGCKKI FGSIAFLPESFDGDPASNTAPLOPELOVEPTLEITTYLYISAWPNSLP
DLSVFONLOVIRGRILHNGAYSLTLOGLISWLGRLSRLRELSGLALIHNTHLCTFVHTVMDOLFERNPH
OALLHTANRPEDCEVGEGLACHOLCARGHCMGPPTOCVNCOSQFLRGOECVEBCRVLOGLPREYVNAHRC
LPHPEQOPONGSVTCGPEADOCVACAHYKDPFCVARCPSGVKPDLSYMPINWKFDEEGACQPCPINC
THSCVDLDDKCGCPAEORASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTWRRLLLOETELVEPL
TPSGAMPNOMRILKETEELRKXVVLGSGAGTYVKGIWIPDGENVKIPVAIKYLRNTSPKANKETLDE
AYMAGVGSPTVSRLLGICLSTVOVLTOLMPYGCCLLDHVBENRGRLGSODLIMCQOIAKMSYLEDR
LVHRDLAARVVLKSPNHVKITDPGLARLDIDETEYHADGKVP IKMALESILRRRFTHQSDVMSYCV
TVELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMIWVKCMIMDSRCRPRFRELVSFBRMA
RDQRFVVLQNEBGLPASPLDSTFYRSLLBEDMDGDLVDAEYVLVPOQGFCCPDPAPAGGMVHHRHRS
STRSGGDLTLGLEPSEEARPSPLASEGSDVFDGLGMGAAGLQSLPTHDPSLQRYSEDPVPL
PSETDGVAPLTCSPDEYVNOPOVRPOPSPREGPLPAARPAANTLERAKTISRKNGVVKDVFVGA
VENPEYLTPOGGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVVI

: Entered [jdelavel 22-Jul-03 6:21]
SB04-59-73-14
MELALICRMGLLIALLPFGAASOVCTGTDMKRLRPLASPETHLDMRLHLYOGCOVVOGNELETYLPFNAS
LSFLDIOEVOQVfivsfw1rpkvaash1eQRLRiVRGTOL.FEDNYALAVLDNGDPLNNTTPVTGASP
GGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNOLATLIDTNRSRACHPCSPMKSGSRCS
MGSESDCOSLTRVCGGCGARCGPLPTDCCHOCAGCTGPKHSDCLALHFNHSGICELHCPALVTY
NTOTFESEMPBERGYTFGASCVTACVYNYLSTVSGCTLYCPLHNOEVTADGTORCEKSGKPCARVCGY
LGMHLEVRVAVTSANIOEPAGCKKI FGSIAFLPESFDGDPASNTAPLOPELOVEPTLEITTYLYISA
WPDLSPLDSVFONLOVIRGRILHNGAYSLTLOGLISWLGRLSRLRELSGLALIHNTHLCTFVHTVMDO
LFRNPOLATHTANRPEDCEVGEGLACHOLCARGHCMGPPTOCVNCOSQFLRGOECVEBCRVLOGLPREY
VNAHRCPLHPEQOPONGSVTCGPEADOCVACAHYKDPFCVARCPSGVKPDLSYMPINWKFDEEGACQ
PCPINTCHSGVDLDDKCGCPAEORASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTWRRLLLOET
ELVEPLTPSGAMPNOMRILKETEELRKXVVLGSGAGTYVKGIWIPDGENVKIPVAIKYLRNTSPKANK
KETLDEAYVWAGVSPVSRLLGICLSTVOVLTOLMPYGCCLLDHVBENRGRLGSODLIMCQOIAKMS
YLEDVRLVHRDLAARVVLKSPNHVKITDPGLARLDIDETEYHADGKVP IKMALESILRRRFTHQSD
VMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMIWVKCMIMDSRCRPRFRELVS
EFSMARDPORFVVLQNEBGLPASPLDSTFYRSLLBEDMDGDLVDAEYVLVPOQGFCCPDPAPAGGMVH
HRHRSSTRSGGDLTLGLEPSEEARPSPLASEGSDVFDGLGMGAAGLQSLPTHDPSLQRYSEDPVPL
PSETDGVAPLTCSPDEYVNOPOVRPOPSPREGPLPAARPAANTLERAKTISRKNGVVKDVFVGA
PARGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVVI
VI

; Entered [jdelaval 22-Jul-03 6:37]
SEQ4-103-117-12
METALACRMGILALILPGAASTOVCTGDMKLRILPASPETHLDMRLHLYOGCOVVOGNLELYLPTNNA
LSFLDIOEVQGVYLIANHVOVQPIQRLRIYRGTOLEFEDNYALAVLDNGDPLNMTNFVSWFVITPKV
OARSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOQLATLIDTNRSRACHPCSPMCKGSRCKGSSSE
DQSLIRIVCAGGACARCKPPLTDCCHBOCAAGCTGPKHSDCLACHFNHSGICELHPALVYVNTDTE
SMNPEGRTTFGASCCTACPYNYLSTDVGSCTLVCPLHNOEYTAEDGTORCEKSKPCARVCYGLMHEH
REVRVTSANTIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEETGYLYISAMPDSL
DLVSQNLQVIRGRIILHNGAVSLTLOGIGISMLGLRSLRELGSGLALIHNNTHLCFVHTVPMQOLFRNPH
OALHTARPEDECVGEGELACHOLCARGHCWGPGPTOCVNCOSQFLRGOCVEECRYLOGLPREYVARHC
LPCHECOPONGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPIWKPDEEGACQPCPINC
THSCVDLDKGCPAEORASPLTISVSAVGIILVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPL
TPSGAMPNOAQMRILKETELRKVKVLSGAFGVYKGIWIPDEENVKIPVAIKVIRENTSPKXKEILDE
AYVAGVGSFYVSRILGICLTSTVOLVTOIMPYGCLLDHVRENRGRIGSODLNMCMQIAKMSYLEYV
LVHRDLAARNVLYKSPNHVKITDIFGLARLLIDDETEYHADGKVPIKMMALESIILRRFTHOSDVSXV
TYWELMTGAKXYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKMMIDSECRPRELIVSEFSMA
RDPORFVUIONEDLGPASPLDSTFYRSLEDDMDGDLVDAEYLVPOQGFCCDPAPAGAGMHHRRSS
STRSGGDLTLLEFSEBEAPRSPPLAPSEGASDVFDGLGMAAKGLOSLPTHDPSPLOKRSDEFTVPL
PSETOGVVAPLTCSPQPEYVNOQDVDPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVXKDVFAFGA
VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 6:39]
SEQ4-103-117-14
METALACRMGILALILPGAASTOVCTGDMKLRILPASPETHLDMRLHLYOGCOVVOGNLELYLPTNNA
LSFLDIOEVQGVYLIANHVOVQPIQRLRIYRGTOLEFEDNYALAVLDNGDPLNMTNFVSWFVITPKV
aashIEQLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOQLATLIDTNRSRACHPCSPMCKGSR
WGSSSDCOSLIRIVCAGGACARCKPPLTDCCHBOCAAGCTGPKHSDCLACHFNHSGICELHPALVY
NTDTPESMNPBERVTFPGASCCTACPYNYLSTDVGSCTLVCPLHNOEYTAEDGTORCEKSKPCARVCY
LGMEHLREVRVTSANTIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVETLEETGYLYISA
WPDLSPLSVFONLOVIRGRIILHNGAVSLTLOGIGISMLGLRSLRELGSGLALIHNNTHLCFVHTVPMQ
LRNPHQALHTARNRPEDECVGEGELACHOLCARGHCWGPGPTOCVNCOSQFLRGOCVEECRYLOGLPREY
VARHRLPCHPECOPONGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPIWKPDEEGACQ
PCPINCTHSCVDLDKGCPAEORASPLTISVSAVGIILVVLGVVFGILIKRQOKIRKYTMRLLOET
ELVEPLTPSGAMPNOAQMRILKETELRKVKVLSGAFGVYKGIWIPDEENVKIPVAIKVIRENTSPKX
KEILDEAYVAGVGSFYVSRILGICLTSTVOLVTOIMPYGCLLDHVRENRGRIGSODLNMCMQIAKMS
YLEEDVRLVHRDLAARNVLYKSPNHVKITDIFGLARLLIDDETEYHADGKVPIKMMALESIILRRFTHOSD
VMSXGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKMMIDSECRPRELIVS
EFSRMRDQRFVUIONEDLGPASPLDSTFYRSLEDDMDGDLVDAEYLVPOQGFCCDPAPAGAGMHH
HRRSSSTRSGGDLTLLEFSEBEAPRSPPLAPSEGASDVFDGLGMAAKGLOSLPTHDPSPLOKRSDE
FTVPLPSETOGVVAPLTCSPQPEYVNOQDVDPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVXKDV
FAFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPI
VI

```
; Entered [jdelaval 22-Jul-03 6:41]
SE04-149-163-12
MELALCRWGLLALLPFGAASSTOVCTGDMKRLPASPEYTHLDMHLHYOGCVOVQGNLETTYLPNTNAS
LSFLQDIQEVQGVYLIHANOVRQVPLQRLRVRGTOULFEDNYALAVLDNGDPLNNTTPTVTGASFGGLREL
QLRSLEILKXGVLIQNPOLCYODTILMKDQYIkanakIgitelINRSRACHPCSPWCKSGSRMGSSSE
DCOSLTRTVAGGACARCKGPLPTDCHEOCAGCTGPHSGDCLACLHFNHSGICELHCALVNTYNTDFFE
SMNPEGRYTFGASCVTACPVNYLSTVSGCTLVCPILNOBVTAEADGTORCEKSKPCARVCGYLGMBHL
REYVAUTSANIOEPAGCKKIFGSLAFIPESFDGPASNTAPLOPEQLOVFTLEITGTYLISAMPSTLP
DLSVPONLOVIRGRILHNGAVSLTLOGISWLGRLSRLSGSLALIHNTHTLCFVHTVMDQLFNRPH
QALLHTRNRPEDCEVCEGLACHQULCARGHCWGPPTOCVNCOSQFLRGQECVEECRVLOGLPREYVNAHC
LPCHPEQOPNGSVTCFGEPAQOCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFPDEBGAQCPPIINC
TPSGCVDLDKCGCPAEORASPLTISAVVGLLVVLGVVFGILIKRQOKIRKXTMRRLLOETELVEPL
TSGAMENQAKOMRILKETELRKXVVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDB
AYMAGVSGPYVSRLLGICLTSTVOLVTOQLMPYGCILDHVARENGRGSLNMCQIAGKMSYLEDR
LVHRDLAARVULVKS PNHVKITDGLARLIDIDETEHADGKVPIMKMALESILRRRFTHOSDVMSCYV
TYWELMTFGAKPTDGIIPAREIPDLLEKGERLPORPITCTIDVYIMVKCMIDSECRPRFRELVSEFSMA
RDPQRFVIVIONEDLGPASPLDSTFYRSLEDDMDGLVDAEYLVPOQGFPCDPAPAGAGVHRRHRS
STRSGGDLTLGLEPSEEAARSPPLASPEGASDVFDGLGMAKGLQS.LPTHDPSPLOQRYSEDPVPL
PSETDGVVALTCSPOPEYVNOQPVRRPOPPSPREGPLPAARPAQTLEERAKTILSPKXGVVXKDVFAFGA
VENPEYLTPOGGAAPQHPPAPFSPAFDNLYYMDODPREGAPSTFKGTATENPEYLGLDVPI
```

```
; Entered [jdelaval 22-Jul-03 6:41]
SE04-149-163-14
MELALCRWGLLALLPFGAASSTOVCTGDMKRLPASPEYTHLDMHLHYOGCVOVQGNLETTYLPNTNAS
LSFLQDIQEVQGVYLIHANOVRQVPLQRLRVRGTOULFEDNYALAVLDNGDPLNNTTPTVTGASFGGLREL
QLRSLEILKXGVLIQNPOLCYODTILMKDQYIkanakIgitelINRSRACHPCSPWCKSGSRMGSSSE
DCOSLTRTVAGGACARCKGPLPTDCHEOCAGCTGPHSGDCLACLHFNHSGICELHCALVNTYNTDFFE
SMNPEGRYTFGASCVTACPVNYLSTVSGCTLVCPILNOBVTAEADGTORCEKSKPCARVCGYLGMBHL
REYVAUTSANIOEPAGCKKIFGSLAFIPESFDGPASNTAPLOPEQLOVFTLEITGTYLISAMPSTLP
DLSVPONLOVIRGRILHNGAVSLTLOGISWLGRLSRLSGSLALIHNTHTLCFVHTVMDQLFNRPH
QALLHTRNRPEDCEVCEGLACHQULCARGHCWGPPTOCVNCOSQFLRGQECVEECRVLOGLPREYVNAHC
LPCHPEQOPNGSVTCFGEPAQOCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFPDEBGAQCPPIINC
TPSGCVDLDKCGCPAEORASPLTISAVVGLLVVLGVVFGILIKRQOKIRKXTMRRLLOETELVEPL
TSGAMENQAKOMRILKETELRKXVVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDB
AYMAGVSGPYVSRLLGICLTSTVOLVTOQLMPYGCILDHVARENGRGSLNMCQIAGKMSYLEDR
LVHRDLAARVULVKS PNHVKITDGLARLIDIDETEHADGKVPIMKMALESILRRRFTHOSDVMSCYV
TYWELMTFGAKPTDGIIPAREIPDLLEKGERLPORPITCTIDVYIMVKCMIDSECRPRFRELVSEFSMA
RDPQRFVIVIONEDLGPASPLDSTFYRSLEDDMDGLVDAEYLVPOQGFPCDPAPAGAGVHRRHRS
STRSGGDLTLGLEPSEEAARSPPLASPEGASDVFDGLGMAKGLQS.LPTHDPSPLOQRYSEDPVPL
PSETDGVVALTCSPOPEYVNOQPVRRPOPPSPREGPLPAARPAQTLEERAKTILSPKXGVVXKDVFAFGA
VENPEYLTPOGGAAPQHPPAPFSPAFDNLYYMDODPREGAPSTFKGTATENPEYLGLDVPI
VI
```

; Entered [jdelaval 22-Jul-03 6:47]
SEQ4-210-224-12
MELALICRMGLLALLPFGAASOVCTGDMKLRLPASPTHLDMLRHLVYGCQVOCNLELTLYPTNAS
LSFLQDIOEVQGVLIJAHNOVROVPLORLIRVGTOLFEEDYALALVLDNGDPLNNTTPTVGTASPGGREL
OLRSLTEILKGGVLIQNNPOLCYODTILMKDI FHKNNQALTLIDTNRSRACHPCSPWCKSGRCMGESSE
DCOSLTRTVACAGCARGKGPlyIkanrkfigtelKHSQCLACHFNHSGICEHLCPALVYNTDTE
SMNPBGRVTFGASCCTACPYNLSTDVSGCTVCPPLHNOEVTAEQTORECKSKPCARVCYGLGMEHL
REVRATVSANIIOEFACCKIFGSLAPLESFDDPASNTPLOPEQLQVETLEITGYLYISAMPDLSL
DLSVFNLOVIRIRILHNGAYSLTLOGIGISMLGRSLRELGSGLALIHNTLHCFVHTVPMQDLFRNPH
QALLHTANRPEDECVGEGLAGHQLCARHGMGPPTQCVCNCSQFLRGQECVEECRVLOGLPREYVNAHCL
LCHPECOBONGSVTCFPGPADQVCACHYKDPFCVARGSPGVKPDLSYMPIMKFPDEBGACOPCTINC
THSCVDLDKGCFAEORASPLTISVSAVVGILVVLGVFGILIKRQOKIRKXTMRLLQETELVEPL
TSGAMPNOAQMRILKETELRKVVLGSGAFVYKGIWIPDEENVKI PVAIKVLRNTSPKANKEILDE
AYMAGVSPYVSRLLGICLTSTVOLVTLQMLPYGCLLDHVRNENGRIGSQDLINWCMQIAKMSYLEDR
LVHRLDAARNVLVKSFNHVKITDGLARLLDIDETEXHADGKVPIMKMALESILRRRFTHOSD
TYWELTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRELIVSEFSRMA
RPORFVVIIONEDIGASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCDPAPAGAGMYHHRHRS
STRSGGDLTLCLEPSEEEAPRSPLASBEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOQRYSE
PSTIDGYVAPLTCSPOPEYVNOQDVPRPQPSREGPLPAARPAAGATLERAKTILSPKNGVVKOVFAFGA
VENPEYLTPOGGAAPQHPHPPAFSPADNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVPI

; Entered [jdelaval 22-Jul-03 6:47]
SEQ4-210-224-14
MELALICRMGLLALLPFGAASOVCTGDMKLRLPASPTHLDMLRHLVYGCQVOCNLELTLYPTNAS
LSFLQDIOEVQGVLIJAHNOVROVPLORLIRVGTOLFEEDYALALVLDNGDPLNNTTPTVGTASPGGREL
OLRSLTEILKGGVLIQNNPOLCYODTILMKDI FHKNNQALTLIDTNRSRACHPCSPWCKSGRCMGESSE
DCOSLTRTVACAGCARGKGPlyIkanrkfigtelKHSQCLACHFNHSGICEHLCPALVYNTDTE
SMNPBGRVTFGASCCTACPYNLSTDVSGCTVCPPLHNOEVTAEQTORECKSKPCARVCYGLGMEHL
REVRATVSANIIOEFACCKIFGSLAPLESFDDPASNTPLOPEQLQVETLEITGYLYISAMPDLSL
DLSVFNLOVIRIRILHNGAYSLTLOGIGISMLGRSLRELGSGLALIHNTLHCFVHTVPMQDLFRNPH
QALLHTANRPEDECVGEGLAGHQLCARHGMGPPTQCVCNCSQFLRGQECVEECRVLOGLPREY
VNAHCLPCHPECOBONGSVTCFPGPADQVCACHYKDPFCVARGSPGVKPDLSYMPIMKFPDEBGACQ
PCPINCHSCVDLDKGCFAEORASPLTISVSAVVGILVVLGVFGILIKRQOKIRKXTMRLLQET
ELVEPLTSGAMPNOAQMRILKETELRKVVLGSGAFVYKGIWIPDEENVKI PVAIKVLRNTSPKAN
KEILDEAYYMAAGVSPYVSRLLGICLTSTVOLVTLQMLPYGCLLDHVRNENGRIGSQDLINWCMQIAKMS
YLEDRVLVHRLDAARNVLVKSFNHVKITDGLARLLDIDETEXHADGKVPIMKMALESILRRRFTHOSD
VMSGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRELIVS
EFSRMAADPQRFVVIIONEDIGASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCDPAPAGAGMYH
HHRHRSSTRSGGDLTLCLEPSEEEAPRSPLASBEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOQRYSE
DPTVPLPSEYDGYVAPLTCSPOPEYVNOQDVPRPQPSREGPLPAARPAAGATLERAKTILSPKNGVVKOV
FAFGAVENPEYLTPOGGAAPQHPHPPAFSPADNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVP
VI

: Entered [jdelaval 22-Jul-03 6:51]
SEQ4-250-264-12
MELALCRWGLLALLPFGAASSTOYCTGTDMKRLRIPASPTHLDMLRHL YQCCQVYQGNLEITYLP TNAS
LSFLQDIOEVGYVLI IAHNOYROYPLQRLRIVRGTOLEFEDNYALALVLDNGDPLNNTTPTVTGASPGGLREB
QRLSLEILKGGVLIQRPOLCYODPTILMKDI FHKNOALALTLLIDTNSRACHPCS PMSKGRCSGSESE
DCOSLTRIVCAGGACRGCPPLTDCCHQCAAGCTGPGHSDCLACIHFHNSGICELHCPALVqy1kanak
figit eLRITFGASCVTACPVNYLSTDVSGCTLVCP LHNQEVTAEDGTORCEKSPCARVYGLGMEHL
REVAVTSANIOEFAGCKKIFGSLAFLPESFGDDPASNTAPLOPELOVFEETLEETIGVLYISAMPDLP
DLVFNOLQVIRGRI LHNQAVSLTLOGIGISWLGRLSRLRELSQFALLHNTHLCFVHTVPMDOFLRPH
QALHTANRPEDECVGEBGLACHQLCARHCWGPGPTQCVNCSQFLRGQECVEBCEVLOGLPREYVNAHC
LPCHPECOPONGSVTCFGEADQVCAHAHYKDPFCVACPSGVKPDLSYMPIMKFPDEEGACQPCPI NC
THSCVLDLDDKCCPAEQRASPLTSIVSAVVGILLVVVLGVFGILLKRQOKIRKXTMRRLQETELVEPL
TPSGAMPNOAQMRLIKETELRKXVLSGAGFTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDE
AYVAGVSPYVSRLLGICTSTVOLVQMLPYGCLDHRNENRGLSGODLNMCMQIAKMSYLEDR
LVHRDLAARNVLVKS PNHVKITDREGARLIDIDEYHADGKVPIMKMALESILRRRFTHQSDVMSYCV
TWELMTFGAPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRFRELVSFSMA
RDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDDKMDLVDAEYLVPOQGFCDPAFAGVHHRHRS
STRSGGDLTLGLSEEEAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPVTPL
PSETDGVAPLTCSPQPEYVNOVDVRPOPSPRREGPLPAARPAAGATLERAKTLSPGKNGVVDV
VENBEYLTPOGGAAPQHPHPAPFAFDNLVYWDODPPERGA PSTFKGTPTAENBEYLGIDVPV1

: Entered [jdelaval 22-Jul-03 6:51]
SEQ4-250-264-14
MELALCRWGLLALLPFGAASSTOYCTGTDMKRLRIPASPTHLDMLRHL YQCCQVYQGNLEITYLP TNAS
LSFLQDIOEVGYVLI IAHNOYROYPLQRLRIVRGTOLEFEDNYALALVLDNGDPLNNTTPTVTGASPGGLREB
QRLSLEILKGGVLIQRPOLCYODPTILMKDI FHKNOALALTLLIDTNSRACHPCS PMSKGRCSGSESE
DCOSLTRIVCAGGACRGCPPLTDCCHQCAAGCTGPGHSDCLACIHFHNSGICELHCPALVfnfvsf
w1vpykvaashleryTFGASCVTACPVNYLSTDVSGCTLVCP LHNQEVTAEDGTORCEKSPCARVYCG
LQMEHLREVRVAVTSANIOEFAGCKKIFGSLAFLPESFGDDPASNTAPLOPELOVFEETLEETIGVLYISA
MPDSLPLDLSVFNOLQVIRGRI LHNQAVSLTLOGIGISWLGRLSRLRELSQFALLHNTHLCFVHTVPMDO
LFRNHQALHTANRPEDECVGEBGLACHQLCARHCWGPGPTQCVNCSQFLRGQECVEBCEVLOGLPREY
VNARHCLECHPECOPONGSVTCFGEADQVCAHAHYKDPFCVACPSGVKPDLSYMPIMKFPDEEGACQ
PCPI NCTHSCVLDLDDKCCPAEQRASPLTSIVSAVVGILLVVVLGVFGILLKRQOKIRKXTMRRLQET
ELVEPLTPSGAMPNOAQMRLIKETELRKXVLSGAGFTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELIDE
AYVAGVSPYVSRLLGICTSTVOLVQMLPYGCLDHRNENRGLSGODLNMCMQIAKMSYLEDR
YLEDVRLVHRDLAARNVLVKS PNHVKITDREGARLIDIDEYHADGKVPIMKMALESILRRRFTHQSD
VMSYGVITWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYMIWVKCMIDSECRPRFRELVS
EFSRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDKMDLVDAEYLVPOQGFCDPAFAGVH
HRHRSSTRSGGDLTLGLSEEEAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPVTPL
PSETDGVAPLTCSPQPEYVNOVDVRPOPSPRREGPLPAARPAAGATLERAKTLSPGKNGVVDV
FAFGAVENBEYLTPOGGAAPQHPHPAPFAFDNLVYWDODPPERGA PSTFKGTPTAENBEYLGIDVP
V1

; Entered [jdelava] 22-Jul-03 7:10]
SEQ4-325-339-12
MEFLADLCRMGLLALLPFGAASQVCTGTDMLKRLPASPETHLDMRLHYOGCOVVOGNLELTYLPTNAS
LSFLADIOEVGVYLAHNOVROVPLORLRIVRGTOLFEEDNYALAVLDNGDPLNNTTPVTGASRGGLREL
OLRLTEILKGGVLIQRNPOLCTODTILMKDIFHKNNOLALTLIDTNRSAHCPCSPMCKGSRMGESSE
DCOSLTRIVAGGCARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHPNHSIGICELCPALVTYNTDTFE
SMNPEGRVYFPGASCVTACPYNTLSTDVGSCTLVCPILHNOBVTAEDEGTORCEKSKPCARVCYGLGMqy1
kansk1gltlelffagckk1fgslafLPESFDGDPASNTAPLOEQLOVEFTLEITGYLYISAMPDPLP
DLSPFONLOVIRGRILHNGAVSLTLOGLISMLGRSLRELISGLALIHNTHLCFVHTVMDQLFENPH
OALLHTANRPEDECVGEGLAHQULCARGHCGMGPPTQCNCOSQFLRGOECVEECRVLOGLPREYVNAHRC
LPCHPCOPONGSVTCGPEADQVCVAHYKDDPPCVARCPGVPDLSTYMPIMKFPDEBEGACPCPINC
THSCVLDLDKGCAPAEORASPLTSIVSAVVGILLVVLGVVPGILIKRQOKIRKTYMRLLQETELVEPL
TBSGAMPNQOMRILKETELKXKVLGSGAFGTVYKGIWLPDGENVKIPVAIKVLRENTSFKANKELIDE
AYVMAVGSPYVSRLLGICTSTVOLVTOLMPYGCILDHVRNENRGRLGSDLLMCMQIAGKMSYLEYV
LVYRDLAARVILVKSPPHVKITDFFGLARLLDIDETEYHAGKVPIMMALESLIRRFTHOSDVMYSYG
TWELTFGAKPYDGIAREIPDLLEKGERLPPOPICTIDVYMIWVKCMIDSECRPRPRELIVSEFSRMA
ROBORFVUIONEDLGSPASPLDSTFYRSILEDMDMDLVDAEYILVPOOGFFCPDPAPAGAGMTHHRRSS
STRSGGDLTLGLPSEEEAPRSPLASBEGAGSVFPGDILGMCAKGLQSLPTHDPSPLOKYSIEDPTVPL
PSETDGYVAAPLTCSPQPEYVNOQDVRPOPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFARFGA
VENPEYLLTPQGAAPQHPHPAPFSPAFDNLVYMDQDPPERGAPSTFKGTPTAENPEYLGIDVPI

; Entered [jdelava] 22-Jul-03 7:10]
SEQ4-325-339-14
MEFLADLCRMGLLALLPFGAASQVCTGTDMLKRLPASPETHLDMRLHYOGCOVVOGNLELTYLPTNAS
LSFLADIOEVGVYLAHNOVROVPLORLRIVRGTOLFEEDNYALAVLDNGDPLNNTTPVTGASRGGLREL
OLRLTEILKGGVLIQRNPOLCTODTILMKDIFHKNNOLALTLIDTNRSAHCPCSPMCKGSRMGESSE
DCOSLTRIVAGGCARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLHPNHSIGICELCPALVTYNTDTFE
SMNPEGRVYFPGASCVTACPYNTLSTDVGSCTLVCPILHNOBVTAEDEGTORCEKSKPCARVCYGLGMq1
ftvafw1rydkvaasnlEEFAGCKK1FGSLAFLPESFDGDPASNTAPLOEQLOVEFTLEITGYLYISA
MPDPLSPFONLOVIRGRILHNGAVSLTLOGLISMLGRSLRELISGLALIHNTHLCFVHTVMDQLFENPH
OALLHTANRPEDECVGEGLAHQULCARGHCGMGPPTQCNCOSQFLRGOECVEECRVLOGLPREY
VNAHRCLPCHPCOPONGSVTCGPEADQVCVAHYKDDPPCVARCPGVPDLSTYMPIMKFPDEBEGACQ
PCPINTCHSCVLDLDKGCAPAEORASPLTSIVSAVVGILLVVLGVVPGILIKRQOKIRKTYMRLLQET
ELVEPLTPSGAMPNQOMRILKETELKXKVLGSGAFGTVYKGIWLPDGENVKIPVAIKVLRENTSFKAN
KEILDEAYVMAVGSPYVSRLLGICTSTVOLVTOLMPYGCILDHVRNENRGRLGSDLLMCMQIAGKMS
YLEEDVLRDLAARVILVKSPPHVKITDFFGLARLLDIDETEYHAGKVPIMMALESLIRRFTHOSD
VMYSYGTWELMTFGAKPYDGIAREIPDLLEKGERLPPOPICTIDVYMIWVKCMIDSECRPRPRELIVS
EFSHMAIDPORFVUIONEDLGSPASPLDSTFYRSILEDMDMDLVDAEYILVPOOGFFCPDPAPAGAGMTH
HRRSSSTRSGGDLTLGLPSEEEAPRSPLASBEGAGSVFPGDILGMCAKGLQSLPTHDPSPLOKYSIE
DPTVPLPSETDGYVAAPLTCSPQPEYVNOQDVRPOPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKD
VFARFGAIVENPEYLLTPQGAAPQHPHPAPFSPAFDNLVYMDQDPPERGAPSTFKGTPTAENPEYLGIDVP
VI

! Entered [jdelaval 22-Jul-03 7:14]
SEQ4-369-383-12
MEIAALCRMGULALALPRGAASOVCTGDMKLRLPASPTHLDMLRHLVGGCVOVGNLELYLPTNAS
LSFLQDIQEVQGYVLIAHNOVROVPLORLRIVRGTOLFEDNYALAVLDNGDPLNNTTPTVGASFGJREL
QLRSLEILKGGVLIQRNPOLCYODTILMKDI FHKNNQALATLIDTNRSRACHPCSPMCKGSRMGESSE
DCOSLRTVACGACGACRKGPLPTDCCHEOCACGCTGPKHSDCLALHFNHSGICELHCPLALVTYNTDTFE
SMNPEGRYFFGASCVTACPVNYLSTVGSCTIVCPILHNOEVTABDGTORCEKSKRCARVCYGLGMqYI
kanakfigitelefagckkiFGSLAFIPESFDDBPASTAPqYIkanakfigiteletgyLYISAMPDILP
DLSVPNLQVIRGRILHNGAVSLTLOGISWLGRLSRELGSGLALIHNTILCFVHTYPMQDLFRNPH
QALLHTANRDEDECVEGLACHQICARGHCHGRCBPPTOCVNCOSQFLRQCECVEBCRVLOGLPREYVNAHC
LPCHPECQOPONGSVTCGPEADQVACAHYKDPFCVARCPSGVKPDLSYMPIMKFPDEBACQPCPINC
THSCVLDLDDKGCAPAEORASPLTSIVSAVVGILVVLGVVFGILIKRQOKIRKYTRRLLOETELVEPL
TSGAMPNQAOMRILKETELRKXYVLGSGAFYVKGIMIPDGENVKIPVAIKYLRNTSPKANKEILDE
AYMAGVSPYVSRLLGICLSTYQVLTOLMPYGCLLDHYRENRGLSGODLWMCQIAGMSYLEDVR
LVHRDLAARNVAVKSPHVKITDGLARLIDIDETEHADGKVPITKMALESILRRRFTHQSDVMSYCV
TYWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPCTIDVYIMVKCMIMDSECRPRFRELVS
RDQKRVVILQNEDLGPASPLDSTFYRSLLLEDDBDMDLVDAEYLVPOQGFPCPDPAAGAGVMHHRSS
STRSGGDDLTLGLEPSEEARSPPLASBEGASDVFDGDLGMGAAGLQSLPTHPDPSPLQRYSEDPVPL
PSETDGVVAULTCSPOPEYVNOQDVRPOQPSPREGPLPAARPAATLERAKTILSPGKGVVXQVFAFGA
VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYMDQDPERGAPSTFKGTPTAENPEYLGLDVPI

! Entered [jdelaval 22-Jul-03 7:14]
SEQ4-369-383-14
MEIAALCRMGULALALPRGAASOVCTGDMKLRLPASPTHLDMLRHLVGGCVOVGNLELYLPTNAS
LSFLQDIQEVQGYVLIAHNOVROVPLORLRIVRGTOLFEDNYALAVLDNGDPLNNTTPTVGASFGJREL
QLRSLEILKGGVLIQRNPOLCYODTILMKDI FHKNNQALATLIDTNRSRACHPCSPMCKGSRMGESSE
DCOSLRTVACGACGACRKGPLPTDCCHEOCACGCTGPKHSDCLALHFNHSGICELHCPLALVTYNTDTFE
SMNPEGRYFFGASCVTACPVNYLSTVGSCTIVCPILHNOEVTABDGTORCEKSKRCARVCYGLGMqYI
kanakfigitelefagckkiFGSLAFIPESFDDBPASTAPqYIkanakfigiteletgyLYISAMPDILP
DLSVPNLQVIRGRILHNGAVSLTLOGISWLGRLSRELGSGLALIHNTILCFVHTYPMQDLFRNPH
QALLHTANRDEDECVEGLACHQICARGHCHGRCBPPTOCVNCOSQFLRQCECVEBCRVLOGLPREYVNAHC
LPCHPECQOPONGSVTCGPEADQVACAHYKDPFCVARCPSGVKPDLSYMPIMKFPDEBACQPCPINC
THSCVLDLDDKGCAPAEORASPLTSIVSAVVGILVVLGVVFGILIKRQOKIRKYTRRLLOETELVEPL
TSGAMPNQAOMRILKETELRKXYVLGSGAFYVKGIMIPDGENVKIPVAIKYLRNTSPKANKEILDE
AYMAGVSPYVSRLLGICLSTYQVLTOLMPYGCLLDHYRENRGLSGODLWMCQIAGMSYLEDVR
LVHRDLAARNVAVKSPHVKITDGLARLIDIDETEHADGKVPITKMALESILRRRFTHQSDVMSYCV
TYWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPCTIDVYIMVKCMIMDSECRPRFRELVS
RDQKRVVILQNEDLGPASPLDSTFYRSLLLEDDBDMDLVDAEYLVPOQGFPCPDPAAGAGVMHHRSS
STRSGGDDLTLGLEPSEEARSPPLASBEGASDVFDGDLGMGAAGLQSLPTHPDPSPLQRYSEDPVPL
PSETDGVVAULTCSPOPEYVNOQDVRPOQPSPREGPLPAARPAATLERAKTILSPGKGVVXQVFAFGA
VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYMDQDPERGAPSTFKGTPTAENPEYLGLDVPI
VI

; Entered [jdelaval 22-Jul-03 7:17]
SEQ4-465-479-12
MELALACRMWGLLALILPGGAASVCTGCTDMKRLRLPASPTHMLRHLVGGCQVVOGNLEITYLPTNAS
LSPLDIOEVOGVYLIANNOVROVPLORLRIVRGTOLFEDNYALAVLDNGDPLNNTTPTVTSRPGSLREL
OLRSLTEILKGGVLIORNPOLCYODTILMKDI FHKNNOLALTIDTNSRACHPCSPMCKGSRCKGSESE
DCOSLTRIVCAGGACRCRKGPLPTCCHEQCAAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTFE
SMNPBGRYTFGASCCTACCPYNYLSTDVGSCTLVCPLNHDEVTABDGTORCEKSKPCARVCYGLMGYI
kanakf1gite1e1EPAGCKI1FGSLAFLPESFDGDPASNTAPLOEQLQVFEFTLEITGYLYISAMPDLP
DLAVFOMLOVIRGRILHNAGVSLTLQGLISMLGRSLRELSGIALIHNHTLCEVHTVPMDOLEFRGYI
kanakf1gite1e1ECVGGELACHOLCARGHCWGPOTOCVNSOFLRGOECVECEGVLOGLPREYVNAHRC
LPCHECOPONGSVTCRGPBADCACAHYKDPFCVARCPSGVKPDLSYMPIMKFPDEBGAOCOPINC
THSCVDLDGCPAEOBASPLTSIVSAVVGILLVVILGVFGILLIKRQOKIRKTYMRLLQETELVEPL
TPSGAMNOMKRLKETELRKVAVLSGAGFTVYKGIWIPDGENVKI1PAIKVIRENTSPRANKELIDE
AYVAGVGSPIVSRLLGICLTSVQVLTQMLPYGCLLDHVRNKGRLGSDLLNMCQI1AKGMSYLEDR
LVHRLAARNVLYVSRPNHVKITDGLARLIDIDEYHADGKVP1KMMALESILRRPRTQSDVMSYGV
TVNELMTFGAKPYDGI1PAREIPDLLEKGERLPOPICTIDVYIMVNCMMIDSECRPFRELVSFSSMA
RDPORFVVI1ONEDLPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGMWHHRSS
STRSGGDLTLGLEPSEEAARSPLABSEAGSDVFDGLMGAKGLQSLPTHDPSPLOQRYSEDPITVPL
PSETDGVVPLTCSPOPEYVNOVDVRFOPSPREGPPLPAARPAAGTLERAKTILSPGKGVVXKDVFAFGGA
VENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGIDVP1

; Entered [jdelaval 22-Jul-03 7:17]
SEQ4-465-479-14
MELALACRMWGLLALILPGGAASVCTGCTDMKRLRLPASPTHMLRHLVGGCQVVOGNLEITYLPTNAS
LSPLDIOEVOGVYLIANNOVROVPLORLRIVRGTOLFEDNYALAVLDNGDPLNNTTPTVTSRPGSLREL
OLRSLTEILKGGVLIORNPOLCYODTILMKDI FHKNNOLALTIDTNSRACHPCSPMCKGSRCKGSESE
DCOSLTRIVCAGGACRCRKGPLPTCCHEQCAAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDTFE
SMNPBGRYTFGASCCTACCPYNYLSTDVGSCTLVCPLNHDEVTABDGTORCEKSKPCARVCYGLMGYI
kanakf1gite1e1EPAGCKI1FGSLAFLPESFDGDPASNTAPLOEQLQVFEFTLEITGYLYISAMPDLP
DLAVFOMLOVIRGRILHNAGVSLTLQGLISMLGRSLRELSGIALIHNHTLCEVHTVPMDOLEFRGYI
kanakf1gite1e1ECVGGELACHOLCARGHCWGPOTOCVNSOFLRGOECVECEGVLOGLPREY
fveafw1rvpkvvaah1aECVGGELACHOLCARGHCWGPOTOCVNSOFLRGOECVECEGVLOGLPREY
VNAHRC1PCHECOPONGSVTCRGPBADCACAHYKDPFCVARCPSGVKPDLSYMPIMKFPDEBGAOC
PCPINCHSCVDLDGCPAEOBASPLTSIVSAVVGILLVVILGVFGILLIKRQOKIRKTYMRLLQET
ELVEPLTPSGAMNOMKRLKETELRKVAVLSGAGFTVYKGIWIPDGENVKI1PAIKVIRENTSPRANK
KEILDEAYVAGVGSPIVSRLLGICLTSVQVLTQMLPYGCLLDHVRNKGRLGSDLLNMCQI1AKGMS
YLDEARLVHRLAARNVLYVSRPNHVKITDGLARLIDIDEYHADGKVP1KMMALESILRRPRTQSD
VMSYGVTVNELMTFGAKPYDGI1PAREIPDLLEKGERLPOPICTIDVYIMVNCMMIDSECRPFRELVS
EFSMARDPORFVVI1ONEDLPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGMWH
HHRSSSTRSGGDLTLGLEPSEEAARSPLABSEAGSDVFDGLMGAKGLQSLPTHDPSPLOQRYSE
DPTVPLPSETDGVVPLTCSPOPEYVNOVDVRFOPSPREGPPLPAARPAAGTLERAKTILSPGKGVVXK
FAFGGAIVENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGIDVP
V1

: Entered [jdelaval 22-Jul-03 7:20]
SEQ4-579-593-12
MELALCRWGLLALLPRGAASVCTGTDMKRLRPAPEETHLDMRLRHLYGCCQVYQGNLEITYLPTRAS
LSFLDODIEVQGYVLIHANQVRQVPLQRLRIVRGTOLEFEDNYALAVLNGDPLNNTTPVTGASPGGLREL
QURSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLALTLLIDTNRSRACHPCSPMGKGRCWGESSE
DCOSLTRIVCAGGACARCGPLPTDCHECCAGCTGPKHSDCLACLHFHNSGICELHCPALVTYNTDPE
SMNPEGVYTGASCVTACGPYNYLSTDVGSCTLVCPILHNOEVTAEJGTORCEKSGPCARVCYGMGYI
kanef:igitelefagckkifgsislaflpesfpgdpasntaPlQPEOLOVFETLEITGYLYISAMPDILP
DLVFNQIVIRGRILHNGAYSILTQIGIGISWLGRLSRLRELSGLALIHNTHLCFVHTVPMDOLEFRNH
OALLHTANRPEDECVBEGGLACHQLCARHCWGPPTQCVCNSQFLRGQECVEECRVLOGLPREYVNAHRC
LPCHPECOPONGSVTCFGEADQVCAAHYKDPFCVACPGYIkanef:igiteLPDEGACQPCPIPC
THSCVDLDDKCGCPAQRASPLTSIVSAVVGILLVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPL
TPSGAMPQAOAMRIKETELRKVYLGSAGFTVYKGIWIPGENVKIPIVAIKVIRENTSPRANKELDE
AYMAGVSGPYVSRLLGICLTSTVOLVTOIMPYGCILDHVRENRGLSGODLLNMCQIAKMSYLEDVR
LVHRDLAARNVLVKSBNHVKITDPGLARLLDIDETEHADGKVPILKMMALESIARRRFTTHOSDVMASYC
TWELMTGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYMIWKCMMIDSECRPRFRELVSERVA
RDPORFVVIQNEIDLGPASPLDSTFYRSLLEDMDMDLVDAEYLVVPOQGFCDPAAPGAGVHHRRSS
STRSGGDLTLGLEPSEEBAPRSPLAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVVL
PSETDGYVAPLTCSPQPEYVNOQDVRRPOPSPREGPLPAARPAATLEBAKTLISGKNGVVDVAFGGA
VENPEYVLTQGGAAQPHPPPAFSPAFDNLVYWDODPBERGAPSTTFKGTPTAENPEYLGLDVPI

: Entered [jdelaval 22-Jul-03 7:20]
SEQ4-579-593-14
MELALCRWGLLALLPRGAASVCTGTDMKRLRPAPEETHLDMRLRHLYGCCQVYQGNLEITYLPTRAS
LSFLDODIEVQGYVLIHANQVRQVPLQRLRIVRGTOLEFEDNYALAVLNGDPLNNTTPVTGASPGGLREL
QURSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLALTLLIDTNRSRACHPCSPMGKGRCWGESSE
DCOSLTRIVCAGGACARCGPLPTDCHECCAGCTGPKHSDCLACLHFHNSGICELHCPALVTYNTDPE
SMNPEGVYTGASCVTACGPYNYLSTDVGSCTLVCPILHNOEVTAEJGTORCEKSGPCARVCYGMGYI
kanef:igitelefagckkifgsislaflpesfpgdpasntaPlQPEOLOVFETLEITGYLYISAMPDILP
DLVFNQIVIRGRILHNGAYSILTQIGIGISWLGRLSRLRELSGLALIHNTHLCFVHTVPMDOLEFRNH
OALLHTANRPEDECVBEGGLACHQLCARHCWGPPTQCVCNSQFLRGQECVEECRVLOGLPREYVNAHRC
LPCHPECOPONGSVTCFGEADQVCAAHYKDPFCVACPGYIkanef:igiteLPDEGACQPCPIPC
THSCVDLDDKCGCPAQRASPLTSIVSAVVGILLVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPL
TPSGAMPQAOAMRIKETELRKVYLGSAGFTVYKGIWIPGENVKIPIVAIKVIRENTSPRANKELDE
AYMAGVSGPYVSRLLGICLTSTVOLVTOIMPYGCILDHVRENRGLSGODLLNMCQIAKMSYLEDVR
LVHRDLAARNVLVKSBNHVKITDPGLARLLDIDETEHADGKVPILKMMALESIARRRFTTHOSD
VMSYGVATWELMTGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYMIWKCMMIDSECRPRFRELVS
EFSRNARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDMDMDLVDAEYLVVPOQGFCDPAAPGAGVHH
HHRSSSTRSGGDLTLGLEPSEEBAPRSPLAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSE
DPVVLPSSETDGYVAPLTCSPQPEYVNOQDVRRPOPSPREGPLPAARPAATLEBAKTLISGKNGVVDV
FAFGAVENPEYLTQGGAAQPHPPPAFSPAFDNLVYWDODPBERGAPSTTFKGTPTAENPEYLGLDVPI
V1

; Entered [jdelaval 22-Jul-03 7:26]
SEQ4-632-652-12
MELAAICRMGLLALLPBGAASTOVCTGTMKULRPASPETHLDMLEHLVOCQOVOCNLELYLPTNAS
LSFLDIOIEVOGVYLIAHNOVROVPLORLIRVGTOLPEDNYALAVLDNGDPLNNTTPTVTSASRGGLREL
QOLSLTEILKXGVLIOHNPOLCYODTILMKDIFHKNNQALTLIDTRSRACHPCSPWCKGSRMGESSE
DCOSLTRIVCAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELHPALVYNTDTFE
SMNPEGRYTFGASCVTACVYNYLSTDVGSCTLVCPHNOBVTAEOSTORCEKSKPCARVCYGLMGYI
kanakfigiteIEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEOLQVETLEITGYLYISAMPDILP
DLVSFONLQVIRGRILHNGAYSLTLOGLISMLGRSLRELGLALIHNTHLCFVHTVPMQDLFRNPH
QALLHTANRPEDECVGEGLACHOLCARGHCMGPGTQCVNCSQFLRQCEVEBCRVLOGLPREVYNARHC
LPCHECQOPONGSVTCFGEPADQVACAHYKDPFCVARCPSGVKPDLSYMPWKPFDEEGACQPCPINC
THSCVDLDDNGCAEORASPLTIS1qy1kanakfigiteIYKROOKIRKYTRRLQETELVEPLTPSGAM
PNOAQMRILKETELRKVYVLSGAFGVYKGIWIPDENVKIPVAIKVIRENTSPKANKEILDEAVYVAG
VGSPIVSRLLGICTSTVOLVTOIMPYGCLLDHYRENRGRLSQDILNMCQIAKMGSYLEDVRLVHRDL
AARNVIVKSPNHVKITDQGLARLLIDETEHADGKVP1KMMALESIILRRRFTHOSDWSYGVTVWELM
TEGAKYVDGIPAREIPDLLEKGERLPOPPICITIDVYIMVCKMMALDESECRPRELIVSESRNARDQRF
VVIQNEDLPPASPLDSTFYSLSLEDDMDLVDAEYILVPOQGFCCPDPAPAGAGVHHHRSSSTRSGG
GDLTLGLEPSEEEAPRSPLAPSEBAGSDVFDGDLGMAAKLOSILPTHDSPLQRYSEDDPTVPLPSTTDG
VYAPLTCSPQPEEVNODVAPQPPSPREGPLPARPAGATLERAKTILSPKNGVYKDVFAFGCAVENPEY
LTPQGAAPQPHPRPAFSPAFDNLVYWDQDPPERGAAPPSTFKGPTAENPEYLGIDVPV1

; Entered [jdelaval 22-Jul-03 7:26]
SEQ4-632-652-14
MELAAICRMGLLALLPBGAASTOVCTGTMKULRPASPETHLDMLEHLVOCQOVOCNLELYLPTNAS
LSFLDIOIEVOGVYLIAHNOVROVPLORLIRVGTOLPEDNYALAVLDNGDPLNNTTPTVTSASRGGLREL
QOLSLTEILKXGVLIOHNPOLCYODTILMKDIFHKNNQALTLIDTRSRACHPCSPWCKGSRMGESSE
DCOSLTRIVCAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELHPALVYNTDTFE
SMNPEGRYTFGASCVTACVYNYLSTDVGSCTLVCPHNOBVTAEOSTORCEKSKPCARVCYGLMGYI
kanakfigiteIEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEOLQVETLEITGYLYISAMPDILP
DLVSFONLQVIRGRILHNGAYSLTLOGLISMLGRSLRELGLALIHNTHLCFVHTVPMQDLFRNPH
QALLHTANRPEDECVGEGLACHOLCARGHCMGPGTQCVNCSQFLRQCEVEBCRVLOGLPREVYNARHC
LPCHECQOPONGSVTCFGEPADQVACAHYKDPFCVARCPSGVKPDLSYMPWKPFDEEGACQPCPINC
THSCVDLDDNGCAEORASPLTIS1fnf1vsfw1tydkvaashLEKROOKIRKYTRRLQETELVEPL
TPSGAMPNOAQMRILKETELRKVYVLSGAFGVYKGIWIPDENVKIPVAIKVIRENTSPKANKEILDE
AVYVAGVSPVSRLLGICTSTVOLVTOIMPYGCLLDHYRENRGRLSQDILNMCQIAKMGSYLEDVRLVHRDL
AARNVIVKSPNHVKITDQGLARLLIDETEHADGKVP1KMMALESIILRRRFTHOSDWSYGVTVWELM
TWEIMTFGAKPTDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVCKMMALDESECRPRELIVSESRMA
RDQRFVVIQNEDLPPASPLDSTFYSLSLEDDMDLVDAEYILVPOQGFCCPDPAPAGAGVHHHRSS
STRSGGDLTLGLEPSEEEAPRSPLAPSEBAGSDVFDGDLGMAAKLOSILPTHDSPLQRYSEDDPTVPL
PSETDGVVALTCSQPEEVNODVAPQPPSPREGPLPARPAGATLERAKTILSPKNGVYKDVFAFGCA
VENPEYLTPOGGAAPQPHPRPAFSPAFDNLVYWDQDPPERGAAPPSTFKGPTAENPEYLGIDVPV1

; Entered [jdelaval 22-Jul-03 7:29]
SEQ4-653-12
MELALCRWGLLALLPRGAASVCTGTDMKRLRPAPEHLDMLRHLXGOCVVOGNTLEITYLPTNAS
LSFLDODIEVOGIVLIANOVROVPLORLIRVGTQLFEDNYALALVLDNGDPLNNTTTPVTGASPGGLREB
QLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHNQOLALTLLIDTNRSRACHPCSPMCKSGRCWGESSE
DCOSLTRVACAGGACARCGPLPTDCHEQCAAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFE
SMNPBEGRYTGASCVTACPNYVLSLDVSGCTLVCCPLHNOEYTAEDGTORCEKSCPCARVCYGLMGYI
kanekfi9ite1EPAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFEETLEETGTLYISAMPDSL
DLAVFONLOVIRGRILHNGAVSLTLOGIGISWGLRSLRELGSGLALIHNTHLCFVHTVPMDOLEFRNH
OALLHTANRPEDECVBEGLAGHOLCARGHCWGPPTOCVNCSPFLRGQECVEECRVLOGLPREYVNAHRC
LPCHECOPONGSVTCFGEADQVACAHYKDPPTCVARCSGVKPDLSYMPIMKFPDEBGAQPCPINC
THSCVDLDDKGCPEAQRASPLTISVAVVGLLVVLGVFGILQYIkanekfi9ite1PLTPSGAMPN
QAQMRILKETELRKVKVLSGAFVTVYKGIWIPDENVKIPVAIKVLRNTSPKANKSILDEAYMAGV
SPYVSRLLGICLTSTVOLVTOIMPYGCCLDHVRENRGLSQDILNMCQIAGKMSYLEVRLVHRDLAA
RNVLVKSPNHYKITDPLGLARLIDIDEYHADGGKVPICKMMALESILRRRFTHOSDMSYGVTWELMTF
GAKYDGIIPAREIPDILEKGERLPOPICTIDVYIMVKCMIIDSECRPRFRELVSFSRMAKDPQRTV
IONEDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGD
LTLEBPESEAPRSPPLAPSEGAGSDVFDGDLGMAAGLOSLPTHDPSPLORYSEDPVPLPSETDGYV
APLTCSPQPEYVNOPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFVAFGAVENPEYLT
POGGAAPQHPHPAPAFSPADNLXYWDQDPPERGAAPRSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 7:29]
SEQ4-653-14
MELALCRWGLLALLPRGAASVCTGTDMKRLRPAPEHLDMLRHLXGOCVVOGNTLEITYLPTNAS
LSFLDODIEVOGIVLIANOVROVPLORLIRVGTQLFEDNYALALVLDNGDPLNNTTTPVTGASPGGLREB
QLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHNQOLALTLLIDTNRSRACHPCSPMCKSGRCWGESSE
DCOSLTRVACAGGACARCGPLPTDCHEQCAAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFE
SMNPBEGRYTGASCVTACPNYVLSLDVSGCTLVCCPLHNOEYTAEDGTORCEKSCPCARVCYGLMGYI
kanekfi9ite1EPAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFEETLEETGTLYISAMPDSL
DLAVFONLOVIRGRILHNGAVSLTLOGIGISWGLRSLRELGSGLALIHNTHLCFVHTVPMDOLEFRNH
OALLHTANRPEDECVBEGLAGHOLCARGHCWGPPTOCVNCSPFLRGQECVEECRVLOGLPREYVNAHRC
LPCHECOPONGSVTCFGEADQVACAHYKDPPTCVARCSGVKPDLSYMPIMKFPDEBGAQPCPINC
THSCVDLDDKGCPEAQRASPLTISVAVVGLLVVLGVFGILQYIkanekfi9ite1PLTPSGAMPN
QAQMRILKETELRKVKVLSGAFVTVYKGIWIPDENVKIPVAIKVLRNTSPKANKSILDEAYMAGV
SPYVSRLLGICLTSTVOLVTOIMPYGCCLDHVRENRGLSQDILNMCQIAGKMSYLEVRLVHRDLAA
RNVLVKSPNHYKITDPLGLARLIDIDEYHADGGKVPICKMMALESILRRRFTHOSDMSYGVTWELMTF
GAKYDGIIPAREIPDILEKGERLPOPICTIDVYIMVKCMIIDSECRPRFRELVSFSRMAKDPQRTV
IONEDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGD
LTLEBPESEAPRSPPLAPSEGAGSDVFDGDLGMAAGLOSLPTHDPSPLORYSEDPVPLPSETDGYV
APLTCSPQPEYVNOPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFVAFGAVENPEYLT
POGGAAPQHPHPAPAFSPADNLXYWDQDPPERGAAPRSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 7:33]
SEQ4-695-709-12
MELALCRWGLLALLPFGAASVCTGTDMKLRLPASPETHLDMLRHLVYOCQOVVQGNLELYLPTNAS
LSFLQDIOEVGVYLIHNOVROVPLQRLRIVRGTOLPEDNYALAVLNDGDLNNTTPTYGASPGGLRE
QLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOALATLITDNRBRACHPCSPMCKGSRCKGSESE
DCOSLRTTVCAGGACRCKGPLPTDCHCOCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTE
SMNPBGRYTFGASCVTACPNYVLTSDVGSCTLVCPLNHOEYTAEDGTORCEKSPCARVCGYGLMGY1
kanakf1gite1EFAGCKKIFGSLAFLESPFDGDASNAPLQPEQLQVFETLEETGYLYISAMPDSL
DLVSFQNLQVIRGRILHNHGAVALTLOGIGISWLGRLSRLRELSGLALIHNNHLCFVHTVPMDOLEFRNH
QALHTANRPREDECEVGEGLACHOLCARGHCWGPPTOCVNCOSQPLRGOECVECECVLOGLPREYVNAHRC
LPCHPECOPONGSVTCFPEADOCVACHYKDPFCVARGSPGVKPDLSYMPIMKRPDEBAGCOPCPINC
THSCVDLDDKCGPAEQRASPLTSIVSAVVGILLVVLGVFGILLIKRQOKIRKTYMRLLQETELVEPL
TPSGAMNQAOQMRILKEQYIKANAKF1gite1TYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDE
AYVAVGSPYVSRLLGICLITSVOLVTLQMPYGCCLDHVRENRLGSDQLNMCQIAKMSYLEDV
LVNRDLAARNVLVKSPPNHVKITDFGLARLLIDETEHADGKVP1KMMALESILRRFTHOSDVMSYGV
TVMELMTFGAPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWKCMMIDSECRPRELIVSEFSMA
RDPQRFVYIQNEDLGPASPLDSTFYRSLLEDMDMDLVADEYLVPOQGFCDPDAFGAGVHHHRHS
STRGGGDLTLGLPSEEEAPRSPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEBDTPVL
PSETDGVVAPLTCSPQPEYVNOQDVRRPOPSPSREGPLPAARPAAGATLERAKTILSPGKNGVADVAFGGA
VENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPI

; Entered [jdelaval 22-Jul-03 7:33]
SEQ4-695-709-14
MELALCRWGLLALLPFGAASVCTGTDMKLRLPASPETHLDMLRHLVYOCQOVVQGNLELYLPTNAS
LSFLQDIOEVGVYLIHNOVROVPLQRLRIVRGTOLPEDNYALAVLNDGDLNNTTPTYGASPGGLRE
QLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOALATLITDNRBRACHPCSPMCKGSRCKGSESE
DCOSLRTTVCAGGACRCKGPLPTDCHCOCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTE
SMNPBGRYTFGASCVTACPNYVLTSDVGSCTLVCPLNHOEYTAEDGTORCEKSPCARVCGYGLMGY1
kanakf1gite1EFAGCKKIFGSLAFLESPFDGDASNAPLQPEQLQVFETLEETGYLYISAMPDSL
DLVSFQNLQVIRGRILHNHGAVALTLOGIGISWLGRLSRLRELSGLALIHNNHLCFVHTVPMDOLEFRNH
QALHTANRPREDECEVGEGLACHOLCARGHCWGPPTOCVNCOSQPLRGOECVECECVLOGLPREYVNAHRC
LPCHPECOPONGSVTCFPEADOCVACHYKDPFCVARGSPGVKPDLSYMPIMKRPDEBAGCOPCPINC
THSCVDLDDKCGPAEQRASPLTSIVSAVVGILLVVLGVFGILLIKRQOKIRKTYMRLLQETELVEPL
TPSGAMNQAOQMRILKEQYIKANAKF1gite1TYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDE
AYVAVGSPYVSRLLGICLITSVOLVTLQMPYGCCLDHVRENRLGSDQLNMCQIAKMSYLEDV
LVNRDLAARNVLVKSPPNHVKITDFGLARLLIDETEHADGKVP1KMMALESILRRFTHOSDVMSYGV
TVMELMTFGAPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWKCMMIDSECRPRELIVSEFSMA
RDPQRFVYIQNEDLGPASPLDSTFYRSLLEDMDMDLVADEYLVPOQGFCDPDAFGAGVHHHRHS
STRGGGDLTLGLPSEEEAPRSPLAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEBDTPVL
PSETDGVVAPLTCSPQPEYVNOQDVRRPOPSPSREGPLPAARPAAGATLERAKTILSPGKNGVADVAFGGA
VENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPI
V1

: Entered [jdelaval 22-Jul-03 7:35]
SEQ4-710-730-12
MELALCRWGLLLALLPGMASTQVCTGTMKRLRLPASPETHLDMLRHLYOGCQVYQNLBELTYLPTNAS
LSFLDDIOEVQGYVLIHNOVQVPLQRLRIVRGTOLEFEDNTALAVLDGDDPLNNTTPVTGASPGGLREL
QRLSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNQALTLIDTNRACGPGSPMGSGRCWGESSE
DCOSLTRVCAGGACARCKGPLPTDCCHOCAGCTGPKHSDCLACLHPHNSGICELHCPALVTYNTDTE
SMNPEGRYTFGASCVTACPVNYLSTDVSGCTLVCPHNOEVTAEQTORCEKCSKPCARVCYGLGMVYI
kanakfigitelefagckki fgsilaflpesfpgdpsnnta plopelovfettlbeittgyltysampdsilp
dlsvfontovirgrlthngavsltlqglgtsmlglsrlsrlsgslal ihnthlcvhtvpmoqlfrnph
oalhtanrpedecvgeglacqolcarghcmwpgptqvcncsoflgqecveecrvloglpreyvnarhc
lpchecopongsvtcfgeadqvacahykdppfcvaccpsgvkpxdlsymptmkpdeegacqpcpinc
ths cvdlldkkgpaeoraspltsiysavvgillvvlgvfglilkrrookirkymrrlloetelvepl
tsgampnqomrriiketelarkvylgsagfgyikanakfigite1virentsprankeildeavymag
vgs pyvsrlilgictstvolvtolmptygclldhvrengrlsgsodlmmcmoiaakmsylledvrlvhrdl
aarnvlvkspmhvkittidfglarlldidetevhadgkxpi kmmalesilrrrptthosdvmasyctywelm
tfgakpydgi pparel pdllekerlpoppi ctidvymimwkcmmidsecprrerelvsesrwardpork
vviqnedlgsppapldstfyrslleddmkgdlvaeeylvpoqgfcppdapagagwvhhrrhssstrsgg
gdltlglepsseeearspplapsegagsdvfdgdlmgaaqglsqslpthdpsplorysedptvpl
vvalptcspopeyvnoqdvrrpoppspregplpaarpagatlbraaktlsbkngvvdvfa fgcgavempey
ltpqgaapqphppafsfafdnlytwdodpbergappstfkgtptaenpeylgldvpvi

: Entered [jdelaval 22-Jul-03 7:35]
SEQ4-710-730-14
MELALCRWGLLLALLPGMASTQVCTGTMKRLRLPASPETHLDMLRHLYOGCQVYQNLBELTYLPTNAS
LSFLDDIOEVQGYVLIHNOVQVPLQRLRIVRGTOLEFEDNTALAVLDGDDPLNNTTPVTGASPGGLREL
QRLSLTEILKGGVLIQRNPOLCYODTILMKDI FHKNNQALTLIDTNRACGPGSPMGSGRCWGESSE
DCOSLTRVCAGGACARCKGPLPTDCCHOCAGCTGPKHSDCLACLHPHNSGICELHCPALVTYNTDTE
SMNPEGRYTFGASCVTACPVNYLSTDVSGCTLVCPHNOEVTAEQTORCEKCSKPCARVCYGLGMVYI
kanakfigitelefagckki fgsilaflpesfpgdpsnnta plopelovfettlbeittgyltysampdsilp
dlsvfontovirgrlthngavsltlqglgtsmlglsrlsrlsgslal ihnthlcvhtvpmoqlfrnph
oalhtanrpedecvgeglacqolcarghcmwpgptqvcncsoflgqecveecrvloglpreyvnarhc
lpchecopongsvtcfgeadqvacahykdppfcvaccpsgvkpxdlsymptmkpdeegacqpcpinc
ths cvdlldkkgpaeoraspltsiysavvgillvvlgvfglilkrrookirkymrrlloetelvepl
tsgampnqomrriiketelarkvylgsagfgyikanakfigite1virentsprankeildeavymag
vgs pyvsrlilgictstvolvtolmptygclldhvrengrlsgsodlmmcmoiaakmsylledvrlvhrdl
aarnvlvkspmhvkittidfglarlldidetevhadgkxpi kmmalesilrrrptthosdvmasyctywelm
tfgakpydgi pparel pdllekerlpoppi ctidvymimwkcmmidsecprrerelvsesrwardpork
vviqnedlgsppapldstfyrslleddmkgdlvaeeylvpoqgfcppdapagagwvhhrrhssstrsgg
gdltlglepsseeearspplapsegagsdvfdgdlmgaaqglsqslpthdpsplorysedptvpl
vvalptcspopeyvnoqdvrrpoppspregplpaarpagatlbraaktlsbkngvvdvfa fgcgavempey
ltpqgaapqphppafsfafdnlytwdodpbergappstfkgtptaenpeylgldvpvi


```
; Sequence 4, Application US/09806703A
; GENERAL INFORMATION:
; APPLICANT: Steinaa, Lucilla
; APPLICANT: Mouritsen, Soren
; APPLICANT: Gautam, Anand
; APPLICANT: Dalm, Iben
; APPLICANT: Haaning, Jesper
; APPLICANT: Leach, Dana
; APPLICANT: Nielsen, Klaus
; APPLICANT: Karlsson, Gunilla
; APPLICANT: Rasmussen, Peter
; TITLE OF INVENTION: Novel Methods for Therapeutic Vaccination
; FILE REFERENCE: 3631-0109P
; CURRENT APPLICATION NUMBER: US/09/806,703A
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/DK99/00525
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: DK 1998 01261
; PRIOR FILING DATE: 1998-10-05
; PRIOR APPLICATION NUMBER: US 60/105,011
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-806-703A--4
; METALICRWGLLLALPGCAASTOVCTGIDMKRLRLPASPETHLDMRLHYOGCOVVOGNLELYL
; PPNASISFLQDIOEOGVVYLIAHNOVROVPIORLRIVRGTOLPEDNYALAVLNGDPIANTTPTVT
; GASPGRLRLQRLSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNQALTLIDNRSRACHPC
; SPWCKSRWGSSEDCQSLTRVCGAGCARCKGPLPTDCHEQCAAGCTGPHSDCLACLAHFNH
; SGI CEIACHPALVTANTDESMNPGRYTFGASCVTACPYNYLSTDVGSCTLVPCPLHNOVTAE
; DGTQCEKSKPCARVCYGLGMEHLREVAITSANIQEFAGCKIFGSLAFLEPSFDGDPASNTA
; PIQPOLQVFEETLEETGLYISAMPDSL.PDLSVFONLOVIRGRILHNGAYSLTLOGLISWGL
; RSLRELGSGLAI.IHNNTHL.CFVHTVMDOL.FRNPQALHTARPEDECVGEGILACHOL.CARGHC
; MGQPTQCNCQSLRAGQCECECRVLOGLPREYVNAHRL.PCHPEQCPQNGSVTCEPPEADOCV
; ACAHYKDPFCVACRCSGVKPDISTYPMKFPDEBACQPCPTNCHSCVDLDDKCCPAEQRASV
; LITSIVSAVGLILVAVLGVVFGILIKRQOKIKRYTMRLLOETELVEPLTPSGAMPNOQMRL
; KETELRKVVLGSGAFGLVYKGIWIPDGQVKIPVAIKVIRENTSPKANKELIDEAYVMAVGSP
; YVSRLLGICLTSTVOLVTOIMPYGLIDDEYRENRGLGSODLNMCMQIAKMGSLYEDVRLVRD
; LAARVVLVSPNHVKITDGLARLIDIDETEVADGQKPIKMALESILARRFTHOSDVMYSGV
; TWELMTGAKRYDCIPABEIPDLLEKGERLPOPICTIDVYIMYKCMWIDSECRPRELYSE
; FSRNARDPQRFVYIOMEDLGPASPLDSTFYSRLLEDDMDLDAEYVLVPOQGFCCPDPAAG
; GNVHRRHRSSTRSGGGLTLGLEPSEEA.PRSPLASEGAGSDVFDGDLGMAAGLOSLPTHD
; PSLQRYSEDPVPLPSETDGVVAPLTCSPQPEYVNOQVVRPQPSPREGPLPARPAGATLERA
; KLSGKNGCVNDFAFGAVENPEYLTPOGAA.PQHPPPAFSPA.FDNLVYWDQDPPERGADPS
; TKGPTAENPEYLGIDVPVI
```

```
; Sequence 12, Application US/09806703A
; GENERAL INFORMATION:
; APPLICANT: Steinaa, Lucilla
; APPLICANT: Mouritsen, Soren
; APPLICANT: Gautam, Anand
; APPLICANT: Dalm, Iben
; APPLICANT: Haaning, Jesper
; APPLICANT: Leach, Dana
; APPLICANT: Nielsen, Klaus
; APPLICANT: Karlsson, Gunilla
; APPLICANT: Rasmussen, Peter
; TITLE OF INVENTION: Novel Methods for Therapeutic Vaccination
; FILE REFERENCE: 3631-0109P
; CURRENT APPLICATION NUMBER: US/09/806,703A
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/DK99/00525
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: DK 1998 01261
; PRIOR FILING DATE: 1998-10-05
; PRIOR APPLICATION NUMBER: US 60/105,011
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 12
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Clostridium tetani
; US-09-806-703A-12
; QYTKANSKFTGLTELI
```

```
; Sequence 14, Application US/09806703A
; GENERAL INFORMATION:
; APPLICANT: Steinaa, Lucilla
; APPLICANT: Mouritsen, Soren
; APPLICANT: Gautam, Anand
; APPLICANT: Dalum, Iben
; APPLICANT: Haaning, Jesper
; APPLICANT: Leach, Dana
; APPLICANT: Nielsen, Klaus
; APPLICANT: Karlsson, Gunilla
; APPLICANT: Rasmussen, Peter
; TITLE OF INVENTION: Novel Methods for Therapeutic Vaccination
; FILE REFERENCE: 3631-0109P
; CURRENT APPLICATION NUMBER: US/09/806,703A
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/DK99/00525
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: DK 1998 01261
; PRIOR FILING DATE: 1998-10-05
; PRIOR APPLICATION NUMBER: US 60/105,011
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-09-806-703A-14
FNNFTVSPWLKVPKVSASHLEI
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - protein search, using sw model

July 22, 2003, 08:05:49 ; Search time 126.375 Seconds
(without alignments)
2046.206 Million cell updates/sec

US-09-806-703A-4

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

US-09-806-703A-4
6812
1 MELALCRWGLLALPPGA.....TFKGTFAENPEYIGLDVIV 1255

Result No.	Score	Query Match	Length	DB	ID	Description
1	6295	92.4	1259	6	O18735	O18735 canis fam1
2	3171	46.6	1209	11	O9GX70	O9GX70 rattus norv
3	3142	46.1	1210	11	O9EP98	O9EP98 mus musculu
4	2751	40.4	1165	13	O9VH40	O9VH40 xiphophorus
5	2729.5	40.1	1137	13	O9K6F6	O9K6F6 gallus gall
6	2326	34.1	1328	13	P79754	P79754 fugu rubrip
7	2044.5	30.0	1433	5	O9BIH9	O9BIH9 anopheles g
8	1874.5	27.5	419	4	O9UK79	O9UK79 homo sapien
9	1739	25.2	729	15	O86712	O86712 mus musculu
10	1720	25.2	729	15	O86712	O86712 mus musculu
11	1718	24.9	567	15	O86714	O86714 avian rous-
12	1697.5	24.3	412	4	O8WV0	O8WV0 homo sapien
13	1653.5	24.3	962	15	O64895	O64895 avian eryth
14	1645	24.1	545	15	O85468	O85468 avian eryth
15	1533.5	22.5	655	11	O9WVFS	O9WVFS mus musculu
16	1517.5	22.3	643	11	O9ERV6	O9ERV6 mus musculu

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

Result No.	Score	Query Match	Length	DB	ID	Description
1	6295	92.4	1259	6	O18735	O18735 canis fam1
2	3171	46.6	1209	11	O9GX70	O9GX70 rattus norv
3	3142	46.1	1210	11	O9EP98	O9EP98 mus musculu
4	2751	40.4	1165	13	O9VH40	O9VH40 xiphophorus
5	2729.5	40.1	1137	13	O9K6F6	O9K6F6 gallus gall
6	2326	34.1	1328	13	P79754	P79754 fugu rubrip
7	2044.5	30.0	1433	5	O9BIH9	O9BIH9 anopheles g
8	1874.5	27.5	419	4	O9UK79	O9UK79 homo sapien
9	1739	25.2	729	15	O86712	O86712 mus musculu
10	1720	25.2	729	15	O86712	O86712 mus musculu
11	1718	24.9	567	15	O86714	O86714 avian rous-
12	1697.5	24.3	412	4	O8WV0	O8WV0 homo sapien
13	1653.5	24.3	962	15	O64895	O64895 avian eryth
14	1645	24.1	545	15	O85468	O85468 avian eryth
15	1533.5	22.5	655	11	O9WVFS	O9WVFS mus musculu
16	1517.5	22.3	643	11	O9ERV6	O9ERV6 mus musculu

ery Match 92.4%; Score 6295; DB 6; Length 1259;
 st Local Similarity 92.1%; Pred. No. 0;
 tches 1160; Conservative 39; Mismatches 55; Indels 6; Gaps 2;

1 MELALCRWGLLALPPGAASVQVCTGDMKRLRASPETHLDMRLHYOGCQVQGNL 60
 1 MELAMCRWGLLALPPGAASVQVCTGDMKRLRASPETHLDMRLHYOGCQVQGNL 60
 61 ELTYLPTNASSLFLDIOBOVQYVLIANNVQVPLORLIRVGTQLFEDNALVLDNG 120
 61 ELTYLPTNASSLFLDIOBOVQYVLIANNVQVPLORLIRVGTQLFEDNALVLDNG 120
 121 DPLNNTTAVTASPGGLRELRLSTELLKGVLIQRNPOLCYDPTIMKDIHKNQOLA 180
 121 DPLNNTTAVTASPGGLRELRLSTELLKGVLIQRNPOLCYDPTIMKDIHKNQOLA 180
 121 DPLNNTTAVTASPGGLRELRLSTELLKGVLIQRNPOLCYDPTIMKDIHKNQOLA 180
 121 DPLNNTTAVTASPGGLRELRLSTELLKGVLIQRNPOLCYDPTIMKDIHKNQOLA 180
 181 LTLIDTRNRSRACHPCSPKCRSGESSEDCOSLTRVNCVGCRCRGLPTDCHEQC 240
 181 LTLIDTRNRSRACHPCSPKCRSGESSEDCOSLTRVNCVGCRCRGLPTDCHEQC 240
 241 AAGCTGPKHSDCLALPHNHSIGELHCPALVTYNTDTFESMPNREGRTGASCTACP 300
 241 AAGCTGPKHSDCLALPHNHSIGELHCPALVTYNTDTFESMPNREGRTGASCTACP 300
 241 AAGCTGPKHSDCLALPHNHSIGELHCPALVTYNTDTFESMPNREGRTGASCTACP 300
 301 YNYLSTDVGSCTLVCPAHNOEVAEDGTORCEKSKPCARVQYGLGMEHLREVAVTAN 360
 301 YNYLSTDVGSCTLVCPAHNOEVAEDGTORCEKSKPCARVQYGLGMEHLREVAVTAN 360
 301 YNYLSTDVGSCTLVCPAHNOEVAEDGTORCEKSKPCARVQYGLGMEHLREVAVTAN 360
 361 IOEFACCKKIFGSLALPESFSDPDASNTAPLOPELOVETLEITGYLYISAMPDLP 420
 361 IOEFACCKKIFGSLALPESFSDPDASNTAPLOPELOVETLEITGYLYISAMPDLP 420
 361 IOEFACCKKIFGSLALPESFSDPDASNTAPLOPELOVETLEITGYLYISAMPDLP 420
 421 DLSVFNLOVIGRIIANGAYSLTLOGLISWLGSLRLSGLALIHNTLCFVHTV 480
 421 DLSVFNLOVIGRIIANGAYSLTLOGLISWLGSLRLSGLALIHNTLCFVHTV 480
 421 DLSVFNLOVIGRIIANGAYSLTLOGLISWLGSLRLSGLALIHNTLCFVHTV 480
 481 PWDOLFERNHQAOLHTANPEDECEGELACIOLCARHCGKGPFOVCNCSQFLRGQC 540
 481 PWDOLFERNHQAOLHTANPEDECEGELACIOLCARHCGKGPFOVCNCSQFLRGQC 540
 481 PWDOLFERNHQAOLHTANPEDECEGELACIOLCARHCGKGPFOVCNCSQFLRGQC 540
 541 VEECEVLOGLPREYVNAHRLCPHCECOFONGSVTCFGEADOCVACAHYKDPFCVANC 600
 541 VEECEVLOGLPREYVNAHRLCPHCECOFONGSVTCFGEADOCVACAHYKDPFCVANC 600
 541 VEECEVLOGLPREYVNAHRLCPHCECOFONGSVTCFGEADOCVACAHYKDPFCVANC 600
 540 VEECEVLOGLPREYVNAHRLCPHCECOFONGSVTCFGEADOCVACAHYKDPFCVANC 600
 601 PSQVPLDLSWPIKMPDEBAGACOPPCINCHSCVDLDKGPAPORASPLTISYAVNG 660
 601 PSQVPLDLSWPIKMPDEBAGACOPPCINCHSCVDLDKGPAPORASPLTISYAVNG 660
 601 PSQVPLDLSWPIKMPDEBAGACOPPCINCHSCVDLDKGPAPORASPLTISYAVNG 660
 600 PSQVPLDLSWPIKMPDEBAGACOPPCINCHSCVDLDKGPAPORASPLTISYAVNG 660
 661 ILVVLGVVGLIKRROQIKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 720
 661 ILVVLGVVGLIKRROQIKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 720
 661 ILVVLGVVGLIKRROQIKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 720
 660 ILVVLGVVGLIKRROQIKRYTMRLLQETELVEPLTPSGAMPNOAQMRILKETEL 720
 721 RKVKVLSGAGFYVYKGIWIPDGENVKIPVAIKVRENTSPKAKELIDEXYVAGVGP 780
 721 RKVKVLSGAGFYVYKGIWIPDGENVKIPVAIKVRENTSPKAKELIDEXYVAGVGP 780
 721 RKVKVLSGAGFYVYKGIWIPDGENVKIPVAIKVRENTSPKAKELIDEXYVAGVGP 780
 720 RKVKVLSGAGFYVYKGIWIPDGENVKIPVAIKVRENTSPKAKELIDEXYVAGVGP 780
 781 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVRERHGRGLGSDLLMWCQIAKGSYLEDDR 840
 781 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVRERHGRGLGSDLLMWCQIAKGSYLEDDR 840
 781 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVRERHGRGLGSDLLMWCQIAKGSYLEDDR 840
 780 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVRERHGRGLGSDLLMWCQIAKGSYLEDDR 840
 841 LVHRDLAANVAVKSNHAKITDGLARLIDETETHADGAVPIKMALESILRRFT 900
 841 LVHRDLAANVAVKSNHAKITDGLARLIDETETHADGAVPIKMALESILRRFT 900
 841 LVHRDLAANVAVKSNHAKITDGLARLIDETETHADGAVPIKMALESILRRFT 900
 840 LVHRDLAANVAVKSNHAKITDGLARLIDETETHADGAVPIKMALESILRRFT 900
 901 HOSDWSVAVTWEMLTFGAKPYDGIPIAREIDPLEKGERLPPPICTIDVYIMKXCM 960
 901 HOSDWSVAVTWEMLTFGAKPYDGIPIAREIDPLEKGERLPPPICTIDVYIMKXCM 960
 901 HOSDWSVAVTWEMLTFGAKPYDGIPIAREIDPLEKGERLPPPICTIDVYIMKXCM 960
 900 HOSDWSVAVTWEMLTFGAKPYDGIPIAREIDPLEKGERLPPPICTIDVYIMKXCM 960
 961 IDSECRPFRELVSFESRARDPQFVYQNEIDLGPASLDSTFYRSLLEDDMDGLVDA 1020
 961 IDSECRPFRELVSFESRARDPQFVYQNEIDLGPASLDSTFYRSLLEDDMDGLVDA 1020
 961 IDSECRPFRELVSFESRARDPQFVYQNEIDLGPASLDSTFYRSLLEDDMDGLVDA 1020
 960 IDSECRPFRELVSFESRARDPQFVYQNEIDLGPASLDSTFYRSLLEDDMDGLVDA 1020
 1021 EBYLVPOGFCPCDPAPAGAGVHHRHSSSTRSGGDLTLGLBFSSEBAPRSLAPSG 1080

Db 1020 EBYLVPOGFCPCDPAPAGAGVHHRHSSSTRSGGDLTLGLBFSSEBAPRSLAPSG 1079
 Qy 1081 AGSDVDGDLGMAAGLQSLPTHDSPLOVSEDPVPLPSESTDGVAPLTSPOEYV 1140
 Db 1080 AGSDVDGDLGMAAGLQSLPTHDSPLOVSEDPVPLPSESTDGVAPLTSPOEYV 1139
 Qy 1141 NOPDVRPOPSPREGPLPARPAGATLER-----AKTSPKXGVKVDVAFGAVENPE 1195
 Db 1140 NOPDVRPOPSPREGPLPARPAGATLER-----AKTSPKXGVKVDVAFGAVENPE 1195
 Qy 1196 YLPQGAAPQHPHPPAPFADNLVYMDQDPPEPARGPSTKGTPTAENPEYGLDVEV 1255
 Db 1200 YLPQGAAPQHPHPPAPFADNLVYMDQDPPEPARGPSTKGTPTAENPEYGLDVEV 1259

RESULT 2

ID 090X70 PRELIMINARY; PRT: 1209 AA.
 AC 090X70;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=90258888; PubMed=2342466;
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA Earp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RL encoded by an alternatively spliced transcript in normal rat tissue.";
 RL Mol. Cell. Biol. 10:2973-2982 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Gutridge K., Dawson T.L., Earp H.S.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RL EMBL: M37394; AAF14008.1; -
 DR HSSP; P11362; IFCK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR Prodom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Receptor; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6C1B7773 CRC64;

Query Match 46.6%; Score 3171; DB 11; Length 1209;
 Best Local Similarity 50.3%; Pred. No. 26-230;
 Matches 642; Conservative 167; Mismatches 354; Indels 114; Gaps 26;

3 LAALCRWGLLALPPGA-ASTVQCTGDMKRLRASPETHLDMRLHYOGCQVQGNL 61

```

Db 15 LAALCAAG-----GALEKKVCQGSNRLTQLTFEDHFLSLORMNCEVULNIE 66
Qy 62 LTYLPTNMSLFLQDIQEVQGYVLIHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
Db 67 IYVORNYDLSTFKTIQEVAGYVLIHNOVRIPLNQLIIRGNALXYNTALAVLSN-- 124
Qy 122 PLNNTPTVPGASPGGIREQLRSLTEILKGVLIQENPOLCVQDTLMDIFHKNNQL 181
Db 125 -----YGNKKTGRLRPLMRLQELIIGAVFSPNNPILCNWETIOMBDIV-QDVFLN 175
Qy 182 TLIDITNRS-RAGHPCSPMKSGSRCESSSEDCOSLTRVCAGGCA-RCKGPLPTDCCHQ 239
Db 176 MSMDOVRHLTGCPKDCPSCPNGSCWGRGENCQKLTIKICAOQCSRCRCGRSPDCCHQ 235
Qy 240 CAAGCTGPRHSDCLCLAHFNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCYTAC 299
Db 236 CAAGCTGPRHSDCLCHFRDPRBATCKDTCPMLLVPPTYQMDVNEGRKYSFGATCVKXC 295
Qy 300 PNYVLTSTVSGCTVCPNLHNOEVLADGQROCEKCKPCARVCYGLGMEHLEVRVATSA 359
Db 236 PNNYVLTSTVSGCTVCPNLHNOEVLADGQROCEKCKPCARVCYGLGMEHLEVRVATSA 354
Qy 360 NIOEPAGCKKIFGSLAFLEPESFGDPASNTAPLOPEQLOVFETLEITGYLIASAMPDL 419
Db 355 NIKHFKYCTAIJSGDLHLIPVAFKGSFTRTPPLDRELEILKTVEKITGFLIQAMPENW 414
Qy 420 PDLSPFQNLQVIRGRILHNGAVSLTQGLISMLGRSLRELSGLALIHNNHTLCEVAT 479
Db 415 TDLHAFENLEIRGRTHQHQGFSLAVVGLINTSLGRSLKEISDGVIIISGRNLCTANT 474
Qy 460 VPMQDLFRNPHQALHTARPEDECCEGLACHQCARHCWGPGTQCVNCSQFLRGOE 539
Db 475 INMKLFGTPNOKTKIMNNRAEKDCATNHCNPLCSSSGSGPEPTDVCSONSRGHE 534
Qy 540 CVEECRVLOGLREYVNARHCLPCHPECOPONGSVTCFPEADOCACAHYDPPRCVVR 599
Db 535 CVDKKNILEGEPREREVNESECICQCHPECPOMNITCTRGDNCKCHAHYDPPRCVVR 594
Qy 600 CPSPGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDLDKCCPAPQORASP-LTSIVASV 658
Db 595 CPSPGIMGENNTL-VWKFADANNVCHLCHANCTYGCAGPLKGC-QQPEGPKIPSIATGI 651
Qy 659 VGLILVVLGVVFGI-LIKRROOKIRKYMRRLDQETELVEPLTPSGAMPNOQRILKE 717
Db 652 VGLILVVLGVVFGI-LIKRROOKIRKYMRRLDQETELVEPLTPSGAMPNOQRILKE 710
Qy 718 TELRKVYLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKXKEILDEAYVMAGV 777
Db 711 TEFKKIKVLSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKXKEILDEAYVMAGV 770
Qy 778 GSPVSVRLIGICLTSTVOLVTQMLPFGCLLDHVRENRGLSGODLNMCMQIAKGMSTLE 837
Db 771 DNPVYCRLLIGICLTSTVOLVTQMLPFGCLLDHVRENRGLSGODLNMCMQIAKGMSTLE 830
Qy 838 DVRLVHRDLAANVIVKSPNHKIKITDFGLARLLDDETYHADGKVPKIMWALSILRR 897
Db 831 DVRLVHRDLAANVIVKSPNHKIKITDFGLARLLDDETYHADGKVPKIMWALSILRR 890
Qy 898 RFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPPCTIDVYMIWVK 957
Db 891 IYTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPPCTIDVYMIWVK 950
Qy 958 CMVIDSECRPRFRELIVSEFSRMAPDQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDMD 1016
Db 951 CMVIDSECRPRFRELIVSEFSRMAPDQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDMD 1010
Qy 1017 LVDAEYVLPVPOGFCPPDAPAGAGMHHRHSSSTRSGGGLTLGLEFSEEARSPPLA 1076
Db 1011 VVDADYEYLPVPOGFCPPDAPAGAGMHHRHSSSTRSGGGLTLGLEFSEEARSPPLA 1036
Qy 1077 PSEGASGVDFDGLCMGAAGKGLSLPTHDPSPLORYSEDPVLPSPET--DGVAVPLTCS 1134
Db 1037 SSLGANSN-----SSYVACINRGSCRVKEDATLQRYSSDPTSVLTEDNIDTFL----- 1086

```

```

Qy 1135 POPEYVNPDPVAPQPPSPREGPLPAARPAATLERAKTISPGKGVYKQVFAFGAVENP 1194
Db 1087 PVPEYVNIQ-SVPRKPAQSVQNPVYHNQPLHP-----APGRDLHYQN--PSHNAVSNP 1135
Qy 1195 EYL-TRQGAAPQHPHPPAPFAFDNLTYWDQ-----DP-----PENGAPPTSF 1237
Db 1136 EYLTNTAQ-----PTCLSSGFDSSALMTQKSHQMSLDNDYOODFFPKAKPWGIF 1186
Qy 1238 KGTPEANPEYLGIDVP 1254
Db 1187 KG-PTNENAEYLRVAP 1202

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
ID Q9EP98, 01-MAR-2001 (TREMBLrel, 16, Created)
AC Q9EP98, 01-MAR-2001 (TREMBLrel, 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr1 transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Scheel C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maile N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr1 transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG28045.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; Cytochrome b5.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Euk_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```


SEQUENCE 1165 AA; 129614 MW; 7F7E38D8771A7AE CRC64;
Query Match 40.4%; Score 2751; DB 13; Length 1165;
Best Local Similarity 45.9%; Pred. No. 1e-198;
Matches 584; Conservative 163; Mismatches 385; Indels 140; Gaps 29;
QY 1 MELAALCEWGLLALLPFG-AAST---OVCSTDMKRLRLPASPEHLMHLHYOGCOV 55
DB 4 LELLEL-----LILLLLISIGRCSTDPDRKYCOGTSNQMTM---LDNHVYKMKMKWYSGCNV 56
QY 56 VQGNLELYLPTNASTLFLODIQEOGVYLIANNOVROVLRRLRVRTQOLFEEDNYALA 115
DB 57 VLENLEITYOENQDLSFQSIQEOGVYLIANNEVSTIPLVNLRLIRQONLYEGNFILL 116
QY 116 VLDNDPLNNTTPTVGSFGLRELOLRSTELKGVLIQRNPOLCYODTILMKDI FPK 175
DB 117 VMSNOK -NPSP--DIVQVGLKQLOLSLTLTGSGVASHNPILCNVETINMDIYVK 173
QY 176 NNQALTLITNRSAPCHPCSPCKSGSRGWSESSSEDQSLTRTVACAGG-ARCKGPLPTD 234
DB 174 TSNPTMNLIPHAFEROCQKCDPGCVNGSCWAPGPHGQKFTKLCAEQCNRCRCGPKPID 233
QY 235 CCECCAGCTGPKKSDCLACHPHNSGICELHCPALVTYNTDPESMNPREGRTFGS 294
DB 234 CNEHCAGCTGPRATDCLACRDEPDGTCCTCPPKIYDIVSHQVNDNPNIKTFPGA 293
QY 295 CVTACPYNYLSTDVSCITLVCPHNOEVTAEDTORCEKSKPCARCYGLGMEHLREVR 354
DB 294 CVKECPNMYVTE -GACVRSAGMLEVD -ENGKSKCKCDGCPVCCIGIGISLSTNI 351
QY 355 AVTSANIOEPACCKKIFGSLAFIPESFGDPAASNTAPLOPOLQVFETLEETGYLYISA 414
DB 352 AVNSTNISFSNCTKNGDIILNRNSFEGDPHYKICGPMPEHMLTLTKEITGYLYIWM 411
QY 415 WPDLSLPDLSVPONLOVIRGRILHNGAYS-LTIQGLISWLGRLSRLSGSLALHNMH 473
DB 412 MPENNTLSLVFONLEIRGRITTSRGSFVNVQVSHLQWLGSLSLKEVSAAGVILLKNTPO 471
QY 474 LCFVATVPMDOLEFRPHQALHTANRPEDECVGEGLAQHLCARGHGMCPGTQCVCSO 533
DB 472 LRYASTIMRRLFRBEDOSIEVDART-----ENQTCNNESEDCGMPGTMCVSCIAH 524
QY 534 FLRGOECVEECRVLOGLPREYVNAHCLFCHPECPONGSVTCFGEADQVACAHYKDP 593
DB 525 VDRGRCVASCNLLGGEPRFAOVDRGVQCHQECIVQTSITCYGGRGANCKCAHPQGG 584
QY 594 PFCVARCISGVKPDLSYPMIKFPEDEGACORCPINCHTSCVYDLDDKCPAQRASPLTS 653
DB 585 PQCIPTCPHGMIGDGDTL-IMKYADKMGQCCQPCCHQNCCTQGCSPGLSGCRGD-IVSHSL 642
QY 654 IVSAVVGILLVVLGVFGLIKRRQOKIRKTYMRLLOETELVEPLTPSGAMPNOAQR 713
DB 643 AVGLVSGILLITYIYVALLIVLRRRRIRK-RKRTIRLLOEKLEVERLTPSGAPNOAFLR 701
QY 714 ILKETLRKVKVLSGAFGTVYKGIWIDGENVKIPVAILKVLRENTSPYANKIIDEAVY 773
DB 702 ILKETEFKKDRVLSGAFGTVYKGLMNPGENIRIPVAILKVLRENTSPYANKIIDEAVY 761
QY 774 MAGVSPVYSLRGICLTSTYQVLTQMLPYGCLLDHVENRRLSODLLANCMQIAKM 833
DB 762 MASVDHPVCRLLGICLTSAVOLVQMLPEYGLLDYVRHORIRICQWILMNCVOIAKM 821
QY 834 SYLEEDVLRHRLDLARNVLKSPNHVKITDFFGLARLLDIDETEVADGKVPKIMMALES 893
DB 822 NYLEBRHLVHRDLARNVLKKNPNVKITDFFGLARLLDIDETEVADGKVPKIMMALES 881
QY 894 ILRRRFTHQSDVSYGVTVMLMTFGAKPYDGI PAREIPDLLEKEERLPQPICTIDVYM 953
DB 882 ILQWYTTQSDVSYGVTVMLMTFGSKPYDGI PAKELIASVLENERLPQPICTIBVYM 941
QY 954 IMVKCMMDISECRPFRELYSEFSRMARDPQRFVVIYQWEDLGPAPLPDSTFRSLLEDD 1013
DB 942 IILKCMMDIPSSRPFRELYSEFSQARDPSRYLVIQ--NLPSPSDRRLFSRLSSDD 998

QY 1014 MGDLYDAEYLYVPOQFFPCPDPAAGAGVWHHRSSSTRSGGDLTLGLEPSEEBAPRS 1073
DB 999 --DIVDADAEYLL-----RYKRIN-RQGS-----E 1019
QY 1074 PLAPSEGAGSDVFDGDLGMAKGLQSLPTHPSPLOKYSBDPTV-PLPSETDGYAPLT 1132
DB 1020 PCIPRNGH-----FVRENSILRYSIDPTQNALELDLGH----- 1054
QY 1133 CSPPEYVNPDPVRQP-----PSPRE-----GPLP-AARPAGTLERAKTSLSPCKXG 1179
DB 1055 -----EYVNPQSESTSSRLSDIYNPYEDLTQMGVSVLSQEAETNFSRPEYLNQNS 1109
QY 1180 VVKDVFAPFGAVENPEYLTPOGGAAPQHPAPPAFPNLVYWDODPPERGAPSTFGK 1239
DB 1110 L---PLVSSGSDDDPY---QAG-----YQAAF-----LQQTGALTGNGMF 1144
QY 1240 TPTAENPEYGL 1251
DB 1145 LPAENLELYGL 1156
RESULT 5
Q9W6F6 PRELIMINARY; PRT; 1137 AA.
AC Q9W6F6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN ERBB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=HINDBRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT embryonic chick hindbrain".
RL Mol. Cell. Neurosci. 13:237-258 (1999).
DR EMBL: AF121963; AADJ1764.1; --
DR HSSP: P1352; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001368; TNFR_C6.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; Ylf_motif.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 1.
DR Pfam: PF02757; YLP_2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWNM_1.
KW Kinase; Tyrosine-protein kinase.
FT NON_TER
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;
Query Match 40.1%; Score 2729.5; DB 13; Length 1137;
Best Local Similarity 47.3%; Pred. No. 4.1e-197;
Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;
QY 161 LCYQPTILMKDIFHKQNLALTLITNRSRACHPCSPCKSGSRGWSESSSEDQSLTRTVAC 220

Db	LCFADTTHMOI VNRPMASNTLTVPTNKS SGGCRHKSCTG - RCMGPRFENHCOITLTKVC	61
Qy	221 AGGC - ARCKGRLPTDCHCEQCAAGCTGPKHSDDLACLHPNHSGLCELHCPALVTYNTDTF	279
Db	62 AEQCSGRGCGYVNSDCHRECAAGGSGGKDTDTCFACMNFNSGACVTCGPOTPFVNNPTTF	121
Qy	280 ESMRPEBERYFGASCYTAACPNTYSTVGSTLVCPLHNOEVTACEDTQCEKSKRCA	339
Db	122 QLEHNHNKATYTGACVCKKCPNNF - VDSSCVYRACPSCKHEV - EBNGLKMKCPPTDTCR	179
Qy	340 RVCYGLGMEHLREYNAVATVSANTQEPAGCKKI FGSJAPLPESEFDGPASNTAPLOBOLOV	399
Db	180 KACOGIGTGLVSAQTVDSSNIDKFINCTKINGNLIFLVTGHGDPYHTIAAINDEKINI	239
Qy	400 FETLEETIGLYISAMPDPSLPRLSPFQULQVIRGAILNGAUSLTLQGLGISMGLBRLR	459
Db	240 FOYVAEITGYLINI OSMPEMNTD FRVFSNLVTIGRALTSGSLLLKQOGLTSLQFSOLK	299
Qy	460 ELGSGLATIHNHTHLCFVHTVPMWDLFRNPHOALHTNRPDEDCVGEGLACHOLCAAH	519
Db	300 QISAGNIITDNSNLCTYHTWMTSLFSTPSQKYVITHNKAKENCTAGCMCNELCSSDG	359
Qy	520 CMGPGPTCVNCSOFLRGOECVEBECVLOGLFREYVNAHCLCPHCEOP - ONGSVTCRG	578
Db	360 CMGPGPDCLSCKREIRGRTCTIESCNLYDGEFREERANGSVCMEDCPOCEKMEEDMNI TCYG	419
Qy	579 PEADOCVA CAHKDPPFCVARCPSGVKPLSLMPLMKPFDEGAOOPORINCTHSCVOLD	638
Db	420 PEPDCHTCKFHKDGPNCVEKCPDGLQGANSE - IFKIADEBCHPCHPNCTQCRGA	477
Qy	639 DKGC-----PABORASPLTISYAVV - GILLVVVLGVFGLIKRQOKIRKYT	686
Db	478 SHDCIYPTROSTLPQHAR - TPL--IAAGVIGCFIIVIMLTFVAVYRRKSIR - KRA	533
Qy	687 MRLLLOETELVRLPSPGAMPNOADMRLIKETELKRVLTGSGAGCTGYKGIWIDENV	746
Db	534 LRRFL - ETELVEPLTPSGTAPNOADRILKETELKRVKXLSGAGCTGYKGIWBEGETV	592
Qy	747 KIPVAIKVLRENTSPEKANEILDEAYVVAAGVSPYVSRLLGLCTSTVOLVLOLMPYGL	806
Db	593 KIPVAIKILNETTGGKRAVEFMDALIMASMDHPLVRLGLVCLSPITQLVTLQMPHGL	652
Qy	807 LDHVRBNRGLCSODLLWCMQIAGMSTLEDEVRLVNRDLAARNLVUSPNHYKITDRGL	866
Db	653 LDYVNEHKNISOLLNWCVOIAGMWYLERRLVNRDLAARNLVUSPNHYKITDFGL	712
Qy	867 ARLILIDETEVADGKVPILKMMALESIARRFTIOSDVMSGYVWELMTFGAPYDGI	926
Db	713 ARLLEGEDEVYADGKMPIKMMALCEIHYRKFTHIOSDVMSGYVTIWEIMTGGKPYDGI	772
Qy	927 PAREIPDLLEKGERLPPOPICITIDVYIMVWCMMIDSECRPFRELVSFSHMAEDPORF	986
Db	773 PPREIPDLLEKGERLPPOPICITIDVYVWVWVWCMMIDADSRPFKEITAEFSSHMAEDPORF	832
Qy	987 VTIQNMED - LGPASPLDSTFYRSLLEDDMDGLVDAAEYLVPOOGFPCRPAPAGAGMTH	1044
Db	833 LVIQGDNRKLSPDPSDKFQNLDEEDLEDDMDAAEYLV - OAFNIPPIYTSRTIDS	891
Qy	1046 RHRSSSTRSGGDLTLGLEPSEEEAPRS - PLAP - SEGAGSDVPDDLDLGMGAAGLQSLP	1102
Db	892 NNNQFYVRRGGAABEGV - PMPYRABGCIIPEARVAVGATAEIFEDTCNGTLRQVATL	950
Qy	1103 THDPBPLQRYSEDPTVPLPS-----ETDGYAPLTCSPQREYVQNQPVPRQPSRPG	1155
Db	951 AKEDSSTORYSADPTVFIPERVIRGELDEBDGYTTPMRDKPKXTDYLPAEENFVSRKNG	1011
Qy	1156 PLPAA - RPAAGATILERAKTSLPGKNGVVKOV-----AFGAVENPEYTLPPOGAAR	1200
Db	1011 DLQAVDNPEYHN-----APNGQRYAABEYVNERPLYNTFANTYLENAYTL-----K	1051
Qy	1207 PHPPPAFAPFNLYWDDPPERKA--DPSTFKGTFT-----AENPEYLV	1249

Db	1056	NNLPEKAKKAFDNDPDYNNHSLPFRSTLIQHPDYLOESTKFFYKONGRIRPIVAENPEYL	1114
RESULT	6		
ID	P79754	PRELIMINARY;	PRT; 1328 AA.
AC	P79754;		
DT	01-MAY-1997 (TREMBLrel. 03, Created)		
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Elx8b3.		
DE	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontidae; Takifugu.		
OX	NCBI_TaxID=31033;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99177347; PubMed=10077531;		
RA	Gellner K., Brenner S.;		
RT	"Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu		
RT	rubripes."		
RL	Genome Res. 9:251-258(1999).		
DR	EMBL; AF056116; AAC34391.1; -.		
DR	HSSP; P11362; IFGK.		
DR	InterPro; IPR000494; EGFR_L domain.		
DR	InterPro; IPR000719; Euk_Dkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001245; Tyr_kinase.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; Dkinase; 1.		
DR	Pfam; PF01030; Recep_L domain; 2.		
DR	Prodom; PD000001; Euk_Dkinase; 1.		
DR	SMART; SMO0261; Fu; 3		
DR	SMART; SMO0219; TyrcK; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP_1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
KW	ATP-binding; Transferase.		
SQ	SEQUENCE 1328 AA; 148613 MW; A331039258B647E9 CRC64;		
Query Match	34.1%; Score 2326; DB 13; Length 1328;		
Beet Local Similarity	40.7%; Pred. No.1.6e-166;		
Matches	524; Conservative 154; Mismatches 418; Indels 190; Gaps 31		
QY	9 WGLTALALPP--GAATG---VCTGTDMLRLPASPETHLMDLRHLVQGCQVQVGNLDEL 62		
DB	4 WRLILMCVASRLRAASSQGTQVACPGTNGSLSTGSCGQNVNLMKDRYKGCETIIMGNLEI 63		
QY	63 TYLPNNASLSTLQDIQEVQGVYVLIANQVQVPLQRLRIYVGTQLFENYVALATVINDDP 122		
DB	64 TQIESNMDFSLKTIREVYGVYLIAMNFOEIPLOGLVINGNSLYERRFALSYFLN--- 120		
QY	123 LNNTPVVGASPGGLRELYLSTLEILKGVYLIQNPOLCTQDTLLMDIFHKNNQLALT 182		
DB	121 -----YPKG--PSGLNGGLANLLEITLDGVIQIINNKTILRGPMVYMWDI--RNNDAITE 173		
QY	183 LIDTRRSACHPCSPMCKGSRCHWESSSDQSLTRTVAGGC-ARCKGPLPTDCHEQA 241		
DB	174 IQFNGERGVCH---KSC-GNYCMWGGKQCOILITKTVCAPOCNDRCFGTSPBDCHIECA 229		
QY	242 AGCGPKRHSDDLACLHFHNSGICELHCPALVTYNTDFEENMPNPEGRYTFGASCTACPY 301		
DB	230 AGCGKPLDTPACKLFLFDSGACVPOCQTLIYNKQTFOMETNPVAKYQVSGICVSQCP 289		
QY	302 NYLSTDVGSCTLVGPRLHNQEVTAEDGTOR-CEKSKPCARVCYGLMGEHLREVRVATVSAN 360		
DB	290 HFV-VDGSSCVSVCPDMEV--ERGSGRQCELSGCLPKVCEGTGAE---QRTVDSN 343		
QY	361 IQEFGAGCKKTIGSLAFIPESFDGPRASNTALQLEPQLOVFTLEIITGYLYISAMPDLP 420		
DB	344 IDSFNCTKIQGSLHFLVTGLIGDFKVVAPPLDAKKLEVFRTVEIITDIILNIQSWPKELN 403		

```

QY 421 DLSVQNLQVINGRIILHNGAYSLTLQGLQISWLGLSLRELSGLALIHNNHLCFVHTV 480
DB 404 DLSVFSLSLTIOGRSLFKRSLMVMRIPLTSLGLSLRELSGVSIVISQNALCYHTHTV 463
QY 481 PMDOLFNRPH-QALHTARPEDECVGELAGHQLCARHGCGRPTCCVNCQFARGE 539
DB 464 NMTQLFRSGRVANSLSNRPMABCVADRVCPLCSDSGCGWPGPDQCLSCNNYSRHGT 523
QY 540 CVEECRVLOGLPREYVNAH-CLPCHPECOQNGSVTCFGEADQCVACAHYKDPFCVA 598
DB 524 CVAGCHFNGSIPREFAGLNGVACVACHPECKPOTGKASCCTGPGADBECACTKPRDGYCMS 583
QY 559 RCPSEVKPDLSPYMPKPPDEGACOPCPINCTHSCVDLDDKCPAPRASPSTISAV 658
DB 584 SCPAGVN-DGEXGLIFKFPNREGHCEPCHQNCQCGSGPLNDC--LEAARLTISSGQI 639
QY 659 VGLLVVVLGVVF-----GILIKRQOKIRKYTRRLLOETELVEPLTPSAMNQAO 711
DB 640 TGIAGCPAGLIFCLVLPFLGMLYHGLAIRRQAKRRYLESGESEPELGP-GEKGTKVH 698
QY 712 MRILETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPYAIKVLRENTSPKANKELIDBA 771
DB 699 ARILKPSDLRKIKPLSGSVFTVSKGFWIPBETVXIPVAIKTIQDSSGRQFTETITDL 758
QY 772 YMAVGSPYVSRLLIGICTSTVQVLTQMPYGCCLLDHVENKRGILGSDLLMCMQIAK 831
DB 759 LSMGSLDHPYIVRLIGICPGTCLQVLTQLSHGLSLEHTROHKTSLDPLRLMVCQIAK 818
QY 832 GMSYLEDVRLVARDLAARVLVKSPMHVKITDGLARLDDIETETHAOCGRPIKMMAL 891
DB 819 GMYVEEHLVHVKNLAKRNIILKNDYQVVISDGVADLLYPDKKYVSEYETKPIKMMAL 878
QY 892 ESILRRRFTHQSDVSVYGVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIV 951
DB 879 ESILRRRFTHQSDVSVYGVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIV 938
QY 952 YMIWVKCMWIDSECPREFELVSEFSRMAQDPQRFVIVIONEDLGASPLDSTFFYSRLSD 1011
DB 939 YMWVKCMWIDENIRPTFKELASDFTRMARDPRYLIVMEG-----ED 982
QY 1012 DDMGDLVDAEVLVYQGGFCPPDPAPGAGMHHRRSSSTSGGDLTLGLPSEEPAR 1071
DB 983 SCMGFFL-----RQSSER--GLLEADLEEDDEE-- 1008
QY 1072 RSLPLASEGAGSDVFDGDLGAG--AAKGLQSLPTHDPSPLO-----RYSEDPVY 1118
DB 1009 -----GLDRTATPSLOPSMSSTSPSQINSYMTQRLRYD----- 1044
QY 1119 PLPSETDGVADLTCSPOP-EYVNO-----PDVRRPPSPREGPL--PAAR 1161
DB 1045 -FAVSGGHIQVLPMSPSFVDITRQIMYQSRSLSVRTLPRDSAPFRSSREABLCEDGAQ 1103
QY 1162 PAGALLERAKTISPCKNGVVKVFAFGCAVENPEYITPQCGAAPQHPHPPASPAFDNLY 1221
DB 1104 CAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1222 YMDODPPERGAAPSTFGTPTAENPE 1247
DB 1123 --QQRKSTASPSFKTMAADEBDE 1146

```

RESULT 7

```

Q9BIH9 PRELIMINARY: PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DB Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

```

```

OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ301655; CAC35008.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000345; Cyt_c_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00757; Furin-1ike; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR KW Receptor.
FT NON TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 30.0%; Score 2044.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred. No. 3,2e-145;
Matches 473; Conservative 196; Mismatches 393; Indels 385; Gaps 36;

26 CNGTDKRLPASPENHMLHLVQCGVQNGLETLPTNASSLPDIODIOGVGYVL 85
1 CIGNGKMSVPANREHYKQLDRYNTCYVDNLSLTIYQNTDNLNFIQHLREVGYVL 60
QY 86 IAHNVQVPLQRLRIVRGTOLE--EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140
DB 61 ISLYLDPQVILPRLQIIRGRTTFKLNKMEAVGLFV-----SFSMNTL 104
QY 141 QRLSLTEILKGGVLIQRPQLCYQDTILMKDI-FHKNQDLATLIDTNSRACHPCSPMC 199
DB 105 ELPALRDLIGSGSVGFENNNTLCHKMSINWEEILAPQTSMQYTFNFPSSPERVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSLTRIVCAGGCA--RCKGRLPTDCCHEOCNAACCTGPKHSDCLACLH 257
DB 165 EVG-CMGEGANHCQRPSSKLNCSPPQSGRCFGKPRECHLFCAGGCTGPTGSDCLACKN 223
QY 258 FHNSGICELHCPALVYNTDTFESMNPREGRYTFGASCYACPYNTLSTDVGCTLVCPPL 317
DB 224 FYDDGVCGQECPCPMQYINPTNYFWERNPDGKAYGATCYRKCP-ELLLDNGACVAKCPK 282
QY 318 HNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREYRAVTSANTIOFPACCKIIFGSLAF 377
DB 283 GMAPQNSE-----CVPCKGVCPKTCPEGGIVH-----SDNIGYKQCTIIEGSLLEI 329
QY 378 PESFGDPRASNT-----APLOPQLOVFEETLEITGVLYISAMPDSLPLDSVQNLQ 429
DB 330 DQSFQGFQOVVYNTFSGPRYIKIDPRLVEVFTVKEITGFINIOAHNPFTTLNLYFRNBE 389
QY 430 VIRGRIILHNGAY-SLTLOGISLWGLRLSRLRELSGLALIHNNHLCFVHTVPMDOLFNR 488
DB 390 VVGGRQKLENPLFASVYIVTSLKSLLEKSLKRVNGSIVYILNSDLCFEDIDMSKIKS 449
QY 488 PFOALLHTANRPEDECVGELAGHQLCARHGCHWGPQTPQCVNCSQFLRQDCVEECRVLO 548
DB 450 SDHEVMVQKNRNVATECHBEWGMECSQSGKAGCMGKPGECLECKIVKYKGLDSCK--- 506

```

```

OY 549 GLPREY-VNARHCLPCHPECCFONGSVTCFGEADQCVAAHYKDPFVCARCP----- 601
DB 507 SLPLRYSDSKTCGCHQCKD-----FCYGNEDNCSQWVNGKRFVACFPITTKAM 561
OY 602 -----SGVPRDLSTYMPWKFPD----- 618
DB 562 NGCTCNCHTKVCGRGRPTDITAPDGCISCDKALISDAKIERLMDDESCPDYSDYVL 621
OY 619 -BEG----- 621
DB 622 QEBGPLKOLSGKAVCKKCHPRCKCTGVGFHFOECQECTGKKGECEDECPDIFYANEE 661
OY 622 --ACOPCPINCT-----HSCVDL-----DD-----KCCPAEQ----- 646
DB 682 TRCLPCHQECRCGCHGIDHNECRMLKLEGGPDYATFTVCNSCPASHPIKRPQEA 741
OY 647 -----RASPLTSIVSAVVGILLVVVLGVVFGI---LIKROOKIRKXTM 687
DB 742 GKIGPYCSADSMOSGLIEPOTQKIVMGSMALILLCVFGIAFVLFNRKKNKDAVAM 801
OY 688 RRLLOETELVEPLTPSGAMPNOAOMRIKTELKRVKVLGSGAFGYKGIWIDGENVK 747
DB 802 TMAIAGEDESEPLRPSVNGPLTLRIIIEKAEIRRGVLCMGAFGRVFGKVMPEGSVK 861
OY 748 IPVAIKVLRENTSPPKANKELIDENYVMAVGSFYVSRLLGICLTSTVQLVTLMPYGLL 807
DB 862 IPVAIKVLEWGSSESSKEFLBAYIMASVHPMLIKLAAVCMTHOMMLITOLPFLGCL 921
OY 808 DHVRENGRGLSODLLMCMQIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITDFGLA 867
DB 922 DYVRNNKDKISKALINMSTOARGMAYLEERLVRDLAARVLYKSPNHVKITDFGLA 981
OY 868 RLDDIDETEHADGKVPKIMMALESILRRRFTHOSDVASGVYVWELMTFGAKYDGI 927
DB 982 KLDDSDSEYRAAGKIMRIMALECIRHRVFTSKSDVAFGIIITWELTYGARBYENVP 1041
OY 928 AREPLDLKEGRLEPQPICTIDVYIMVCMIMSECPREPLVSEFSRMAPORFV 987
DB 1042 AKDVDELIEIGHKLPQPDICSLDYCILLCWLDADARPTKOLAETFAKARDBGRYL 1101
OY 968 VIQNEDELASPLDSFTFYRSILEDMDMDLV----- 1018
DB 1102 MI-----PQDKRMRLPSTTNODEKOLIRLAVMAAAAAAGASNVDPSTIA 1152
OY 1019 DAEEYLVPOCGFFCPDPAPAGAGVHRRSSSTRSGGDLTLGLEPSEBEARS----- 1073
DB 1153 ENDEYLOPKTRPSIMLPESA-----VEPS--DEMPKSLRYCK 1188
OY 1074 -PLAP--SEGAGSDVFDGDLGMAKGLQSLRPTHDPSPLORYSEDPVTLPSETDGYVA 1129
DB 1189 DPLKPDDETGHGKEY-----GVGGIR-----LMLPLDEDDYLM 1222
OY 1130 PLTCSQREYVNVQRPQPPSPREGPLPARPAGATLERAKTSLPGKNGVNDVFAFG 1189
DB 1223 P-TCOSO-----NOS-----TPG-----YMDLIGVPA 1243
OY 1190 AVENPEYL-----TPQGAAPQPHPPAPSPAFDNLTYMDQDEPPEKCAPSTFKGT 1240
DB 1244 SVNDPEYLMSTQAIAGLAQOSMG--PHTPP-----PNTNGM 1280
OY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

```

RESULT 8
O9UK79 PRELIMINARY; PRT; 419 AA.
ID O9UK79;
AC O9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.

```

```

GN HER-2.
OC Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
R1 SEQUENCE FROM N.A.
R2 MEDLINE=9415951; PubMed=10485918;
R3 Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
R4 "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
R5 autoinhibitor";
R6 Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
R1 SEQUENCE FROM N.A.
R2 Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
R3 Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
R4 EMBL: AF177761; AAD56009.2;
R5 InterPro: IPR000494; EGFR_L domain.
R6 InterPro: IPR002174; Furin-like.
R7 Pfam: PF00757; Furin-like; 1.
R8 Pfam: PF01030; Recep_L domain; 1.
R9 SMART: SM00261; Fu; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLBE347E2D030C CRC64;

```

```

Query Match 27.5%; Score 1874.5; DB 4; Length 419;
Best Local Similarity 89.6%; Pred. No. 3,6e-133;
Matches 353; Conservative 5; Mismatches 29; Indels 7; Gaps 2;

```

```

OY 1 MELALCRWGLLALLPFGAASVCTGTDKRLIPASPEHLDMLRLHYGCGVVGSL 60
DB 1 MELALCRWGLLALLPFGAASVCTGTDKRLIPASPEHLDMLRLHYGCGVVGSL 60
OY 61 ELTYLPTNASLFLDIOEVQGYVLIANOVQVPLQRLTYRGQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLFLDIOEVQGYVLIANOVQVPLQRLTYRGQLFEDNYALAVLNG 120
OY 121 DPLNNTTPTVTCASPGGLRELQRLTEILKGGVLIQRPOLCYQDTILMKDI FHKNOQA 180
DB 121 DPLNNTTPTVTCASPGGLRELQRLTEILKGGVLIQRPOLCYQDTILMKDI FHKNOQA 180
OY 121 DPLNNTTPTVTCASPGGLRELQRLTEILKGGVLIQRPOLCYQDTILMKDI FHKNOQA 180
DB 121 DPLNNTTPTVTCASPGGLRELQRLTEILKGGVLIQRPOLCYQDTILMKDI FHKNOQA 180
OY 181 LFLIDTNSRACHPSPCKSGRCSGSESDCQSLTRYVACGACRCKRPLPTDCCHEOC 240
DB 181 LFLIDTNSRACHPSPCKSGRCSGSESDCQSLTRYVACGACRCKRPLPTDCCHEOC 240
OY 241 AAGCTGPKXSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRYTFGASCTYAC 300
DB 241 AAGCTGPKXSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRYTFGASCTYAC 300
OY 301 YNTLSTDVSGCTVCGELHNOEYTAEDGTORCEKSKRCARVCYGLGMEHLREYAVTSAN 360
DB 301 YNTLSTDVSGCTVCGELHNOEYTAEDGTORCEKSKRCARVCYGLGMEHLREYAVTSAN 360
OY 361 IOEFACGCKLFGSLAFLESFDDDPASNTAPLQ 394
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSALPLAP 387

```

```

RESULT 9
O8R2X1 PRELIMINARY; PRT; 367 AA.
ID O8R2X1;
AC O8R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
R1 SEQUENCE FROM N.A.
R2 Straube R.;
R3 Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

```

DR EMBL/BC027080.1; AAH27080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 5.1e-123;
 Matches 333; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLEKGERLPQPPIC 948
 DB 1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLEKGERLPQPPIC 60
 QY 949 IDVYIMVCMWIMDSECRPFRELVSFSGHMADPQRFVITQMEDLGPASPLDSTYRSL 1008
 DB 61 IDVYIMVCMWIMDSECRPFRELVSFSGHMADPQRFVITQMEDLGPASPLDSTYRSL 120
 QY 1009 LEDDDMDGLVDAAEYLVPOGFCPPDPAPAGAMVHRRSSSTRSGGDLTLGLPSE 1068
 DB 121 LEDDDMDGLVDAAEYLVPOGFCPPDPAPAGAMVHRRSSSTRSGGDLTLGLPSE 180
 QY 1069 EAPRSPFLAPSEGAQDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYV 1128
 DB 181 EAPRSPFLAPSEGAQDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYV 240
 QY 1129 APLTCSPOREYVNPQVPRQPPSPRSGPLPAAPPAATLERATLSPGKXGVKDVPAFG 1188
 DB 241 APLTCSPOREYVNPQVPRQPPSPRSGPLPAAPPAATLERATLSPGKXGVKDVPAFG 300
 QY 1189 GAVENEYLTPOGAPAPPPAPFPAFNLTYMDDPBERGAPSTFGCTTANPEY 1248
 DB 301 GAVENEYLTPOGAPAPPPAPFPAFNLTYMDDPBERGAPSTFGCTTANPEY 360
 QY 1249 LGLDVFPV 1255
 DB 361 LGLDVFPV 367

RESULT 10
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC 086712;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE POLYPROTEIN.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 NC NCB1_Taxid=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSP; P03322; 1A6S.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR004028; Retro_M.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02813; Retro_M; 1.
 DR ProDom; PDD00001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F614FEED63 CRC64;

Query Match 25.2%; Score 1720; DB 15; Length 729;
 Best Local Similarity 54.8%; Pred. No. 3.9e-121;
 Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PQNGSTTCRPEADQCAACHYDPPFCVACRCSGVKPDLSYPMKMFPEBEGACQCP 628
 DB 141 PEETAPPKTGP--DHCKCAHFIDGPRCVACAGVGENDTL-VWYCAANAVCOLCHP 197
 QY 629 NCTHSCVDDDDKCCPAEQRASPLTSIVSAV--GILLVTVGVFGILIKRQOKIRKRYTM 687
 DB 198 NCTHSCVDDDDKCCPAEQRASPLTSIVSAV--GILLVTVGVFGILIKRQOKIRKRYTM 253
 QY 688 RLLQETELVEPLTPSGAMPNQAOMRILKETELRKVKVLSGAGFVYKGIWIPDGENVK 747
 DB 254 RLLQETELVEPLTPSGAMPNQAOMRILKETELRKVKVLSGAGFVYKGIWIPDGENVK 313
 QY 748 IPAIVTLRENTSPKANKELIDEAAYMAGVGSFYSRLGICTSTVQLVTQMPYGLL 807
 DB 314 IPAIVTLRENTSPKANKELIDEAAYMAGVGSFYSRLGICTSTVQLVTQMPYGLL 373
 QY 808 DHVRENRGLSGODLNMCMQIAKMSYLEVRLVHRDLAARVTVKSPHHVKTDFGLA 867
 DB 374 DHVRENRGLSGODLNMCMQIAKMSYLEVRLVHRDLAARVTVKSPHHVKTDFGLA 433
 QY 868 RLDDIDETEHADGKVPKIMMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 927
 DB 434 KLGADKEKYEHAEGKVPKIMMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 493
 QY 928 AREIPLLEKGERLPQPPICITDVYIMVCMWIMDSECRPFRELVSFSGHMADPQRFV 987
 DB 494 AREIPLLEKGERLPQPPICITDVYIMVCMWIMDSECRPFRELVSFSGHMADPQRFV 553
 QY 988 VIO-NEDLGPASPLDSTFYRSLLEDMDGLVNAEEYLVPOGFCPPCDPAPAGAMVHHR 1046
 DB 554 VIO-NEDLGPASPLDSTFYRSLLEDMDGLVNAEEYLVPOGFCPPCDPAPAGAMVHHR 598
 QY 1047 HRSSSTRSGGDLTLGLPSEEEAPRSPPL-----ASEGASGVFDGDLGMAKGLQSL 1101
 DB 599 -NSPT-----SRTPLLSLSATSNNSATNCID-----RNCQG 631
 QY 1102 PTHDPSPLORYSEDPVPLPSET--DGYAAPLTCSPQPEYVNPQVPRQPPSPRSGPLA 1159
 DB 632 PTHDPSPLORYSEDPVPLPSET--DGYAAPLTCSPQPEYVNPQVPRQPPSPRSGPLA 675
 QY 1160 ARPAGATLERATLSPGKXGVKDVF-----ARCGAVNEPYL 1197
 DB 676 -----TANVQNIYNNISLTAISKLPMSRYONSHSTAVDPEYL 715

RESULT 11
 Q86714 PRELIMINARY; PRT; 567 AA.
 AC 086714;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 NC NCB1_Taxid=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; -
 DR HSP; P11362; 1FGK.
 DR InterPro; IPR000719; Euk_pkinase.

DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; kinase; 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; TyrcK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR Tyrosine-protein kinase.
 KW NON TER
 FT SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25E1 CRC64;
 Query Match 25.2%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 3.8e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCAACHYKPPFCVAPRCPSGVKPDLSYMPIMKRPDEGACOPCPICHTSCVDL 637
 DB 1 GP--DHCKKCAHFIDPHCHVACAPGAVGENDTL--VMKTDANAVCOLCHPCTGCKCP 57
 QY 638 DDKGCPAEQRASPLTSIVSAV--GILLVVLGVVFGILIKRROKIKRYTMRLLQETEL 696
 DB 58 GLEGCP---NSKPTSIAGVVGGLCLVVGGLGILYRR--HIVKRTTLARLLQERL 113
 QY 697 VEPLTPSGAMPNOAQMRILKETELRKVYLGSGAGFTYKGIWIPDGENVKIPVAIKVL 756
 DB 114 VEPLTPSGEAPNOAHLRIKETEFKKVKVLSGAGFTYKGLMIEGEEKVPIPAIKELR 173
 QY 757 ENTSSKAKELIDELYVMAVGSFVYSRLGICLTSTQVLTQMLPYGGLDHPVENRGR 816
 DB 174 EATSSKANEILIDELYVMAVSDNPRVCRLLGICLTSTQVLTQMLPYGGLDHPVENRGR 233
 QY 817 LGSODLLWCMQIAKGMSTLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLDIDETE 876
 DB 234 IGSQVLLWCMQVIAAGMNVLEERLVHRDLAARNLVKTPQHVKITDFGLARLLGADENE 293
 QY 877 YHADGKPIKMALESILRRFTHOSDWSYGVTVWELMTGCAKYDGIIPAREIPDLLE 936
 DB 294 YHAEKGKPIKMALESILHRTYHOSDWSYGVTVWELMTGSKRYDIPASEISVYLE 353
 QY 937 KGERLPPIPICTIDVYMIWKCMIIDSECRPFRELVSFRRMARDQRFVYIQ--NEDG 995
 DB 354 KGERLPPIPICTIDVYMIWKCMIIDADSRPFRELIAFSGMARDPPRYIYIIGDERNA 413
 QY 996 PASPLDSTFYRSLDEDDMDLVDAEYLVPQGFECPPAPAGAGMHHRRSSSTRSG 1055
 DB 414 LPSPLDSKRYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---- 454
 QY 1056 GGDLLGLPESEEARSPPL-----APSEGASDVDFDGLGMAKGLQSLTTHDPSPLQ 1110
 DB 455 -----SRPLSLSLSTSNNSATNCID-----RNGGHPVRBDSFVO 491
 QY 1111 RYSEDPVPLPSET--DGVAVPLTCSPOPEYVNOPVRPOPSPREGPLPAARPAATLE 1168
 DB 492 RYSSDPTGNFLEESIDDGFL-----PABEYVNO--LMPKKS----- 526
 QY 1169 RAKTLSPGNNGVVKOVF-----AFGGAIVENPEYL 1197
 DB 527 ----TAMVQNOIYNNISLTAISKLPMDSRYSQNSHSAVDNPEYL 566

RESULT 12
 O8WYVO PRELIMINARY; PRT; 412 AA.
 AC O8WYVO;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN PP3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OK NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth.";
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF318349; AAL55856.1;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00219; TyrcK; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 8.3e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALSELRRFRTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICT 948
 DB 1 MALSELRRFRTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICT 60
 QY 949 IDVYMIWKCMIIDSECRPFRELVSFRRMARDQRFVYIQ--NEDG 1008
 DB 61 IDVYMIWKCMIIDSECRPFRELVSFRRMARDQRFVYIQ--NEDG 120
 QY 1009 LEDDMGDLVDAEYLVLPQGFECPPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEE 1068
 DB 121 LEDDMGDLVDAEYLVLPQGFECPPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1069 EAPRSPPLASEGASDVDFDGLGMAKGLQSLTTHDPSPLQRYSEDPTVPLPSETDGY 1128
 DB 181 EAPRSPPLASEGASDVDFDGLGMAKGLQSLTTHDPSPLQRYSEDPTVPLPSETDGY 240
 QY 1129 APLTCSPOPEYVNOPVRPOPSPREGPLPAARPAATLEAKTSLSPGNNGVVKOVFAG 1188
 DB 241 APLTCSPOPEYVNOPVRPOPSPREGPLPAARPAATLEAKTSLSPGNNGVVKOVFAG 300
 QY 1189 GAVENPEYLPQGGALSPFTLLPSAQSPTSITGTRTHOSGGLHAPBSKHLRORTOST 1220
 DB 301 GAVENPEYLPQGGALSPFTLLPSAQSPTSITGTRTHOSGGLHAPBSKHLRORTOST 360
 QY 1221 YYWD--QDPPER-----GAPSTFKGTPTAEN 1245
 DB 361 WWMTQCEPREGQVRRSPDVSSGSRGLTSAGIKRWEGPPTTSRGCTHARN 410

RESULT 13
 O64895
 AC O64895; PRELIMINARY; PRT; 962 AA.
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag-V-erb-A, V-erb-B protein.
 GN GAG-V-ERB-A, V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OC NCBI_TaxID=11861;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9020603; PubMed=1969616;
 RA Bruckin A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B.";


```

QY 996 PASPLDSTFYRLLEDDMDGLVDAEEYLPQOGFCPPAPGAGMWHHRHSSSTRSG 1055
DB 414 LSPSTDSKFKYRLMEBEDMDIVADDEVLYPHGCF-----NSPST--- 454
QY 1056 GGDLTITGLEPSEEARSP-----APSEGASDVDFDGLGMAAKGLQSLPHPHDSPLQ 1110
DB 455 -----SRTPLSSLSLATSNNASANCIDRNG-----H----- 481
QY 1111 RYSEDPVLPSETDGVVAPLTCSPQPEYVNOQDVPRQPSPREGPLPARPAGAT-LER 1169
DB 482 -----PVREDFL-----PAPRYNO--LHPKPKSTAMVOIQIYNYISLRISK 523
QY 1170 AKTLSPGKGVVQVDFAFGAVENPEYL 1197
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
ID Q9WVF5;
AC Q9WVF5;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maizle N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/10J, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearseall R.S., Green P.J., Yee D., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maizle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RC MEDLINE=2108560; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schiml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.T., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohzuki S., Hayashizaki Y.;
```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AF124513; AAD44149.1; -.
DR EMBL: AF275366; AAG28047.1; -.
DR EMBL: AF275364; AAG28047.1; JOINED.
DR EMBL: AF275365; AAG28047.1; JOINED.
DR EMBL: AK004944; BAB23688.1; -.
DR EMBL: AK004883; BAB23641.1; -.
DR EMBL: AK004911; BAB23662.1; -.
DR MGI: 95294; Egfr.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR SMART: SM00261; Fur_3.
KM RECEPTOR.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.5%; Score 1533.5; DB 11; Length 655;
Best Local Similarity 44.4%; Pred. No. 4,1e-107;
Matches 284; Conservative 99; Mismatches 233; Indels 23; Gaps 7;

QY 11 LLIALLPQAA--STGYCTGTDMKRLPASPTHLDMLRLYOGCGVQGNLELYLPTN 68
DB 14 LITLCAAGALEEKVKCGTSNRLTQGTFFEDHFLSLQRMVNNCEVLGNLEITYVQRN 73
QY 69 ASLSFLQDIOEVQGYLLAHNOVRYPLRLRYRTGTOLFEDNYALAVDNGPLNNTTP 128
DB 74 YDLSFUKTTIOEAVGYLLALNTVERLPLENLOIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLRELQRLSTELIKGVLIOENPOLCYODTILMKDI-----FKANNQLALTLI 184
DB 125 -YGTNRTGLRELPMRLLOELIGAVFSPNNPILCNMDITQMRIVQNVFNSMMDL--- 180
QY 185 DTRNSRACHPCSPMCKGRSGESSSEDCSLTITVACGGA-RCKGPLPTDCHECCAG 243
DB 181 -QSHPSCKPCDPSCPNGSCMGGEENCKLTICCAQCCHRCGRSPDCCHNCACAG 239
QY 244 CNGPKSDCLACLHFNHSGICELHCPALTYNTDTESNPNEGRTTFGASCTACPRNY 303
DB 240 CTGPRESDCLVQCFQDEATKDTCPMLLYNPTTYQMDVNPBGKYSFGATCVKCKPRNY 299
QY 304 LSTDVSGCTLVCPILHNOEYTAEDGTORCEKSKPCARVCYGLMEHLREVAVTSANIOE 363
DB 300 VVTDHSGCYRACGPDYEV-EBDGIKKCKKCDPCKKVCNGSIGEFKDTLSINATNIXH 358
QY 364 FAGCKKIFGSLAFLEPSFDGDPASNTAPLOEQLOVFETLEETGYLYISAMPDSLPLDS 423
DB 359 FKYCTAISGDHLHLPVAFGDSFTTRPLDPRELILKTYKEITGFLLIQAWPDNMTDLH 418
QY 424 VQONLOVITGRILLNCAVYLLTQGLGISWLGRLSELQSLALIHNNTHLCVHTVPWD 483
DB 419 AFENLEIIRGRTKHQOFSLAVVGLNITSLGRSLKEISDGVILISGNRLCYANTINWK 478
QY 484 OLFRNPQALHTANRPEDECVGEGLACHQLCARGHCWGPQTCNCSQFJRGOCVVE 543
DB 479 KLFGRPNQTKTMNRAEDCCAVNHVCPNLSSBECGWGPBDDCVSCQNVSGRCVEX 538
QY 544 CRYVLOGLPREVYNARHCLCPHECPONGSVTCFGPEADQVACAHYKPPFCVAPCPSG 603
DB 539 CNLIGEPREFEIVNSCICGHECPQLQANNITCTGPGNCTQCAHYIDGPHVKTCPCAG 598
QY 604 VNPDLSTYPIWKFPPDEGACQCPINCTHSCVDLDDKXC 642
DB 599 IGENNTL-VWKYADANNVCHLCHANCTYGACGPGLOGC 636
```

Search completed: July 22, 2003, 08:11:27
 Job time : 133.375 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842
Sequence: 1 MEDIALCRMGLLALLPPGA.....TFKGTPTAENPEVLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6626	96.8	1255	1 A24571	protein-tyrosine k
2	5846.5	85.5	1254	2 148161	p-185 precursor -
3	5846	85.4	1260	1 TYRTNU	protein-tyrosine k
4	3118	45.6	1210	1 GQHUE	epidermal growth f
5	3086	45.1	1210	1 A53183	epidermal growth f
6	3070.5	44.9	1223	1 TVCHLV	epidermal growth f
7	2942.5	43.0	1308	2 A47253	epidermal growth f
8	2661	38.9	1166	1 S06142	protein-tyrosine k
9	2381.5	34.8	1342	2 A36223	kinase-related tra
10	2300.5	33.6	1339	2 UC4387	epidermal growth f
11	1766.5	25.8	698	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYUW	protein-tyrosine k
13	1647	24.1	544	2 S35745	protein-tyrosine k
14	1640	24.0	545	2 S00727	kinase-related tra
15	1626.5	23.8	1330	1 GQFPE	epidermal growth f
16	1623	23.7	540	1 B44776	protein-tyrosine k
17	1621	23.7	540	1 TVFVVB	epidermal growth f
18	1478	21.6	644	2 A63325	epidermal growth f
19	1276	18.6	1323	2 E88257	protein-tyrosine k
20	1276	18.6	1374	2 S70712	protein-tyrosine k
21	1194	17.5	1369	2 S70713	protein-tyrosine k
22	1167	17.1	1717	1 A45558	epidermal growth f
23	1108	16.2	527	2 A42032	epidermal growth f
24	943.5	13.8	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	717	10.5	1363	2 T43220	insulin-like growt
28	703	10.3	1382	1 INHUR	insulin receptor p
29	702	10.3	1383	2 A36080	insulin receptor p

30	701.5	10.3	1372	2 A34157	insulin receptor p
31	691	10.1	1477	2 T18534	protein-tyrosine k
32	683.5	10.0	1300	2 A36502	insulin receptor-r
33	680	9.9	1607	2 T43212	insulin-like growt
34	671	9.8	1268	2 B36502	insulin receptor-r
35	645	9.4	1367	1 IGHUR1	insulin-like growt
36	632.5	9.2	2148	1 A56081	insulin receptor -
37	630	9.2	1390	2 T30346	insulin-like growt
38	625	9.1	1371	2 A33837	insulin receptor -
39	623	9.1	2101	2 S57245	insulin-like growt
40	601	8.8	987	2 A54092	insulin receptor (
41	597.5	8.7	952	2 I50612	protein-tyrosine k
42	597.5	8.7	984	2 A39753	protein-tyrosine k
43	587.5	8.6	977	2 S49004	tyrosine kinase Mp
44	586	8.6	1091	2 S33596	protein-tyrosine k
45	585.5	8.6	976	2 A36355	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N.Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein ei
C.Species: Homo sapiens (man)
C.Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C.Accession: A24571; A25491; A44188; B44188; I59509; I57622
R.Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.
Nature 319, 230-234, 1986
A.Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A.Reference number: A24571; MUID:86118663; PMID:3003577
A.Accession: A24571
A.Molecule type: mRNA
A.Residues: 1-1255 <1AM>
A.Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R.Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A.Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A.Reference number: A25491; MUID:86016729; PMID:2995967
A.Accession: A25491
A.Molecule type: DNA
A.Residues: 737-1031 <SEM>
A.Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R.Coussens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,
Science 230, 1132-1139, 1985
A.Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A.Reference number: A44188; MUID:86070181; PMID:2999974
A.Accession: A44188
A.Molecule type: DNA
A.Residues: 740-910 <COU>
A.Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A.Accession: B44188
A.Molecule type: mRNA
A.Residues: 1-517, 'RALI', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A.Cross-references: GB:M11730; NID:G183986
R.King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A.Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A.Reference number: I59509; MUID:85272597; PMID:2992089
A.Accession: I59509
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 832-909 <REX>
A.Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R.Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A.Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcription
A.Reference number: I57622; MUID:87286898; PMID:3039351
A.Accession: I57622
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AAA58637.1; PID:9553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics:
A:Gene: GDB:ERBB2; NCU; NEU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A>Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F:1-21/Domain: signal sequence #status predicted <Sig>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,253,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.8%; Score 6626; DB 1; Length 1255;
Best Local Similarity 97.1%; Pred. No. 2.1e-263;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
QY 1 MELALCRWGLLALLPAASTOYCTGDMKRLPASPETHLDMRLHYOGQVYQGNL 60
DB 1 MELALCRWGLLALLPAASTOYCTGDMKRLPASPETHLDMRLHYOGQVYQGNL 60
QY 61 ELYTPNASTSFLDIOGVGVYLAHNVQVPLQLRLIVRGTLFEDNYALAVLDNG 120
DB 61 ELYTPNASTSFLDIOGVGVYLAHNVQVPLQLRLIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTASPGGLRELOSLREILKGVLLIQRNPOLCYODITLTKDIFHNKQOLA 180
DB 121 DPLNNTPTVGTASPGGLRELOSLREILKGVLLIQRNPOLCYODITLTKDIFHNKQOLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRITVCAAGCARKCKPLPTDCHEQC 240
DB 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRITVCAAGCARKCKPLPTDCHEQC 240
QY 241 AAGCTGPRKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPRKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPRLNHOEYTAEGTORCEKSPCARVCYGLCMOYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPRLNHOEYTAEGTORCEKSPCARVCYGLCMOYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOQVFTLEETITGLYISAMPDSL 420
DB 361 IOEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOQVFTLEETITGLYISAMPDSL 420
QY 421 DLSVQNLQVTRGRLLANGAYSLLTQIGLISWGLRSRLRELGSGALLIHNHNLCTFVHTV 480
DB 421 DLSVQNLQVTRGRLLANGAYSLLTQIGLISWGLRSRLRELGSGALLIHNHNLCTFVHTV 480
QY 481 PMDOLFRFNFTVSMFLRPKVSASHLECEVGEGLACHQLCARGHGCMGCPQCNCSQF 540
DB 481 PMDOLFRFNFTVSMFLRPKVSASHLECEVGEGLACHQLCARGHGCMGCPQCNCSQF 540
QY 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDP 600
DB 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDP 600
QY 601 FCVARCPSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPAEGRPASPLTSI 660
DB 601 FCVARCPSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPAEGRPASPLTSI 660

DB 595 FCVARCPSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPAEGRPASPLTSI 654
QY 661 VAAVVGILLVWVLGVVFGILIKRROOKIKRTYMRLLQSTELVEPLTPSGAMNQOMR1 720
DB 655 ISAVVGILLVWVLGVVFGILIKRROOKIKRTYMRLLQSTELVEPLTPSGAMNQOMR1 714
QY 721 LKETELRKVKVLSGAFGVYKGIWIPDENYKIPAIKVLRENTSPKANKELDEAYVM 780
DB 715 LKETELRKVKVLSGAFGVYKGIWIPDENYKIPAIKVLRENTSPKANKELDEAYVM 774
QY 781 AGVGSPPYVRLIGICLTSITVOLTQMPYGCLLDHYENRGRIGSODLNMCMQIAKMS 840
DB 775 AGVGSPPYVRLIGICLTSITVOLTQMPYGCLLDHYENRGRIGSODLNMCMQIAKMS 834
QY 841 YLEDVRLVHRDLAARVNLVKSPPHVKITPFGRLALDIDETEHADGKVPIMMALESI 900
DB 835 YLEDVRLVHRDLAARVNLVKSPPHVKITPFGRLALDIDETEHADGKVPIMMALESI 894
QY 901 LRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
DB 895 LRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
QY 961 MYKCMWIDECRPRFRELVSFSSRMARDPQRFVITONEDLGPASPLDSTFYRSLLEDDM 1020
DB 955 MYKCMWIDECRPRFRELVSFSSRMARDPQRFVITONEDLGPASPLDSTFYRSLLEDDM 1014
QY 1021 GDLVDAEELVLPQGFPCFDPAPGACGMVHHNRSSSTSGGDLTLGLBPSSEEA PRSP 1080
DB 1015 GDLVDAEELVLPQGFPCFDPAPGACGMVHHNRSSSTSGGDLTLGLBPSSEEA PRSP 1074
QY 1081 LAPSEAGSDVPDGLGMAAGLQSLPHHDSPLQRYSEDDPVPPLPSETDGVAPLTS 1140
DB 1075 LAPSEAGSDVPDGLGMAAGLQSLPHHDSPLQRYSEDDPVPPLPSETDGVAPLTS 1134
QY 1141 POPEYVNOBPVPOPPSPREGPLPAARAGATLERAKTSLPGNGVGVKDVAFGAVENP 1200
DB 1135 POPEYVNOBPVPOPPSPREGPLPAARAGATLERAKTSLPGNGVGVKDVAFGAVENP 1194
QY 1201 EYLTPOGGAAPOPHPPPASPFNDLYYDODPPERGAPSPFTKGTFTLENBYLGLDVP 1260
DB 1195 EYLTPOGGAAPOPHPPPASPFNDLYYDODPPERGAPSPFTKGTFTLENBYLGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 2
148161
P-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: 148161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: 148161; PMID:94193007; PMID:7908275
A:Accession: 148161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PIDN:BAA03801.1; PID:9747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.5%; Score 5846.5; DB 2; Length 1254;
Best Local Similarity 85.3%; Pred. No. 1.4e-231;
Matches 1080; Conservative 60; Mismatches 109; Indels 17; Gaps 3;

```

QY      1 MELAALCRMGWLLALLPRGAASSTOVCTGTDMKRLPASPEITHLMDLHLYGCGVQGNL 60
DB      1 MELAAMCGWGLLALLSPGASGTQVCTGTDMKRLPASPEITHLMDLHLYGCGVQGNL 60
QY      61 ELTYLPNASTLSFLQDIQEVQGYVLIHANOVROVPLQRLIVRGTOLEFEDNYALAVLDNG 120
DB      61 ELTYLPNASTLSFLQDIQEVQGYVLIHANOVROVPLQRLIVRGTOLEFEDNYALAVLDNR 120
QY      121 DPLNNTPTVPGASPGGLREQLRSLEILKQGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
DB      121 DPLDNTVTTATGRPEGLRELQRLSTLEILKQGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
QY      181 LTLIDTRSRACHPCSPMGKSGRCSWGSSESDCQSLTRVCAAGCARCKPLPTDCCHQC 240
DB      181 LTLIDTRSRACHPCSPMGKSGRCSWGSSESDCQSLTRVCAAGCARCKPLPTDCCHQC 240
QY      241 AAGCTGPKHSDCLACLFHNSGICELCHPALVTNTPTFESMPREGRYTGGASCYACAP 300
DB      241 AAGCTGPKHSDCLACLFHNSGICELCHPALVTNTPTFESMPREGRYTGGASCYACAP 300
QY      301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTORCEKSKPCARVCYGLQMGYIKANSKFIQIT 360
DB      301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTORCEKSKPCARVCYGLQMGYIKANSKFIQIT 360
QY      361 ELEPAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVFTTEITGYLISAMPDLP 420
DB      361 IOEPAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVFTTEITGYLISAMPDLP 420
QY      421 DLSVFOVLQVIRGRIHNGAYSLTLOGLGSIWLGRLSRLREGSLALIHNNTHLCFPHTV 480
DB      421 DLSVFOVLQVIRGRIHNGAYSLTLOGLGSIWLGRLSRLREGSLALIHNNTHLCFPHTV 480
QY      481 PMDOLFRFNFTVFSFMLRVKVSASHL-----EECVBEGSLACHOLCARGHCWGPPTQCV 535
DB      481 PMDOLF-----RNPHQALLHSGNSEEECGKXDFACVYLCAHGHGCMGSGPTQCV 529
QY      536 NCQGFPLRGQCEVECRVLQGLPREYVNAARHCLPCHPEQOPONGSVTGRPPADCCVAAH 595
DB      536 NCQGFPLRGQCEVECRVLQGLPREYVNAARHCLPCHPEQOPONGSVTGRPPADCCVAAH 595
QY      590 YKQSPFCVVARCPGSKPDLSTYMPIMKFPDEEGACOPRINCTHSCVLDLDDGCPABORAS 649
DB      590 YKQSPFCVVARCPGSKPDLSTYMPIMKFPDEEGACOPRINCTHSCVLDLDDGCPABORAS 649
QY      656 PLTSIVASAVVGIILVVLGVVFGILIKRROQKIRKTYMRRLLOETLEVEPLTPSGAMPNQ 715
DB      650 PATSIATVVGILFLVIGVVVGIILIKRROQKIRKTYMRRLLOETLEVEPLTPSGAMPNQ 709
QY      716 AQMRILKETELRKVKVVGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKELD 775
DB      710 AQMRILKETELRKVKVVGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKELD 769
QY      776 EAVYMGVGSPPYVSRILGICLTSTVOLVTOLEMPGCLLDHYRENRGLSGODLLNMCQI 835
DB      770 EAVYMGVGSPPYVSRILGICLTSTVOLVTOLEMPGCLLDHYRENRGLSGODLLNMCQI 829
QY      836 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMM 895
DB      830 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMM 889
QY      896 ALESILRRRTTHOSDWSYGVTTWELTFGAKPYDGIIPAREIPDLLEKGERLEPPICTI 955
DB      890 ALESILRRRTTHOSDWSYGVTTWELTFGAKPYDGIIPAREIPDLLEKGERLEPPICTI 949
QY      956 DVMYIMVCKMMISECPRFRELVSFSPRARDPQRFVVIQNEIDLGASPLDSFFYISL 1015
DB      950 DVMYIMVCKMMISECPRFRELVSFSPRARDPQRFVVIQNEIDLGASPLDSFFYISL 1009
QY      1016 EDDMDGDLVDABEYLVPOQGFCEPDPAFGAGVNHRRSSSTRSGGDTTLGEPSEEE 1075
DB      1010 EDDMDGDLVDABEYLVPOQGFCEPDPAFGAGVNHRRSSSTRSGGDTTLGEPSEEE 1069
QY      1076 APRSPLAPSEGASDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVA 1135

```

```

DB      1070 PPRSPLAPSEGASDVFEELGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVA 1129
QY      1136 PLTCSPOPEYNOBVDVPOPPSPREGPLPAARPGATLERAKTISPGKNGVVDVFAFGG 1195
DB      1130 PLACSPPEYNOBVDVPOPPSPREGPLPAARPGATLERAKTISPGKNGVVDVFAFGG 1189
QY      1196 AVENPEYLTPOGGAAPPHPPPAFSPAFDNLVYMDODPPERGAPESTFKPTAENPEYL 1255
DB      1190 AVENPEYLTPOGGAAPPHPPPAFSPAFDNLVYMDODPPERGAPESTFKPTAENPEYL 1248
QY      1256 GLDVPV 1261
DB      1249 GLDVPV 1254

RESULT 3
TNRNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C/Accession: A24562; A61204
R/Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A/Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A/Reference number: A24562; MUID:86118662; PMID:3945311
A/Accession: A24562
A/Molecule type: mRNA
A/Residues: 1-1260 <BAR>
A/Cross-References: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
R/Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen
Carcinogenesis 12, 1975-1978, 1991
A/Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiazolyl]formamide or N-methyl-N-nitrosourea
A/Reference number: A61204; MUID:92035293; PMID:1682063
A/Accession: A61204
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 637-663 'V', 665-702 <MAS>
A/Note: authors translated the codon GCA for residue 25 as Val
C/Genetics:
A/Gene: neu
C/Superfamily: epidermal growth factor receptor; protein kinase homology
C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
F:1-19/Domin: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domin: transmembrane #status predicted <TMN>
F:733-988/Domin: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: lys #status predicted
F:982,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.4%; Score 5846; DB 1; Length 1260;
Best Local Similarity 85.8%; Pred. No. 1,4e-231;
Matches 1084; Conservative 52; Mismatches 118; Indels 10; Gaps 3;

QY      1 MELAALCRMGWLLALLPRGAASSTOVCTGTDMKRLPASPEITHLMDLHLYGCGVQGNL 60
DB      4 MELAAMCGWGLLALLSPGASGTQVCTGTDMKRLPASPEITHLMDLHLYGCGVQGNL 63
QY      61 ELTYLPNASTLSFLQDIQEVQGYVLIHANOVROVPLQRLIVRGTOLEFEDNYALAVLDNG 120
DB      64 ELTYLPNASTLSFLQDIQEVQGYVLIHANOVROVPLQRLIVRGTOLEFEDNYALAVLDNR 123
QY      121 DPLNNTPTVPGASPGGLREQLRSLEILKQGVLIQRNPOLCYQDTILMKDIFHKNNOLA 179
DB      124 DPLDNTVTTATGRPEGLRELQRLSTLEILKQGVLIQRNPOLCYQDTILMKDIFHKNNOLA 183
QY      180 LTLIDTRSRACHPCSPMGKSGRCSWGSSESDCQSLTRVCAAGCARCKPLPTDCCHQC 239
DB      184 LTLIDTRSRACHPCSPMGKSGRCSWGSSESDCQSLTRVCAAGCARCKPLPTDCCHQC 243

```

```

240 CAAGCTGPKHSDCLACLAHFNHSGICELHCPALVTYNTDTPESMPNREGRTFGASCYTAC 299
244 CAAGCTGKHSDCLACHFNHSGICELHCPALVTYNTDTPESMNPREGRTFGASCYTTC 303
300 PYNALSTDVSGCTLWCPLHNOEVTAEQTCCEKSKPCARVCYGLNQYIKANSKFIGI 359
304 PYNALSTEVGSGCTLWCPLHNOEVTAEQTCCEKSKPCARVCYGLNQYIKANSKFIGI 363
360 TELEFAGCKKIFGSLAFLPESPDGASNTAPLOEQOVPFTLEITGYLYISWPSL 419
364 NVQGFDDGCKKIFGSLAFLPESPDGASNTAPLOEQOVPFTLEITGYLYISWPSL 423
420 PDLVFPQULQVIRGRILHNGAYSLTLQGLISWLSRLSRLSGLALIHNTHLCPYHT 479
424 RDLVFPQULQVIRGRILHNGAYSLTLQGLISWLSRLSRLSGLALIHNTHLCPYHT 483
480 VPMDOLFRNNFTVSWFLRVPKVSASHLE--CVGEGLACHQLCARGHCWGPPTQVCNC 537
484 VPMDOLFRNNFTVSWFLRVPKVSASHLE--CVGEGLACHQLCARGHCWGPPTQVCNC 536
538 SGLRGCCEVCECVLQGLPREYVNAHCLPCHPCOPONGSVTCFGEADQCAAHYK 597
537 SHFLNGCEVCECVLQGLPREYVNAHCLPCHPCOPONGSVTCFGEADQCAAHYK 596
598 DPPECVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDBKGCPCAEQASPL 657
597 DSSSCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDBKGCPCAEQASPL 656
658 TSIVSAVVGILLVAVFGLILKRRQOKIRKYMRLLOETBELVEPLTPSGAMPNOAQ 717
657 TSIVSAVVGILLVAVFGLILKRRQOKIRKYMRLLOETBELVEPLTPSGAMPNOAQ 716
718 MRILKETELRKVKYLGSGAFGVVYKGIWIPGSEANKIVAKIVLEKNTSPKANKILDEA 777
717 MRILKETELRKVKYLGSGAFGVVYKGIWIPGSEANKIVAKIVLEKNTSPKANKILDEA 776
778 YVMAGVSPYVSRLLGICLTSTVOLVTLMPYGLLDHVRNKRGLSGODLIMCMQIAX 837
777 YVMAGVSPYVSRLLGICLTSTVOLVTLMPYGLLDHVRNKRGLSGODLIMCMQIAX 836
838 GMSLLEEDVRLVHRDLAANNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKXAL 897
837 GMSLLEEDVRLVHRDLAANNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKXAL 896
898 ESILRRRFTHOSDWSYGVTWELMTFGAKPYDGIIPAREIPDLKESGRLPQPICTIDV 957
897 ESILRRRFTHOSDWSYGVTWELMTFGAKPYDGIIPAREIPDLKESGRLPQPICTIDV 956
958 YMINVKCMMIDSECRPRELVSERFARMARPORFVIONEDLGSPASPLDSTFYSLLED 1017
957 YMINVKCMMIDSECRPRELVSERFARMARPORFVIONEDLGSPASPLDSTFYSLLED 1016
1018 DDMGDLVDAEYLVPOQGFPCPDPAKGCAMVHHRSSSTRSGGDLTLGLEPSEEPAP 1077
1017 DDMGDLVDAEYLVPOQGFPCPDPAKGCAMVHHRSSSTRSGGDLTLGLEPSEEPAP 1076
1078 RSPPLAPSGAGSDYVDGDLGWAAGKGLSLPTHDSPLQRYSEDPVTLPSSTOGYVAPL 1137
1077 RSPPLAPSGAGSDYVDGDLGWAAGKGLSLPTHDSPLQRYSEDPVTLPSSTOGYVAPL 1136
1138 TCSPQPEYVNOQPPVRRPSPREGPLPAPAPAGATLERAKTLSPKGNVAVDFAFGAV 1197
1137 TCSPQPEYVNOQPPVRRPSPREGPLPAPAPAGATLERAKTLSPKGNVAVDFAFGAV 1196
1198 ENPEYLTPOGGAQAQPHPPAPAFDNLVYWDQPPERGAAPSTFKGTPTAENPEYIGL 1257
1197 ENPEYLTPOGGAQAQPHPPAPAFDNLVYWDQPPERGAAPSTFKGTPTAENPEYIGL 1256
1258 DVPV 1261
1257 DVPV 1260

```

```

RESULT 4
GCHUE
Epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A43615; A23062; A05281; A60143; A
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.;
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <U>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A:Note: the authors translated the codon AAG for residue 540 as Aan
R:ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal g
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <TSH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, F.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termin
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; N
ature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCAMRA', 150-187, 'KSVQNAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-3
, 798-799, 'ND', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF rec
R:Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarky, L.S.; Weber, M.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificatio
A:Reference number: A43615; MUID:84196372; PMID:6328261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, M.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

```

R. Busso, M.W.; Lukae, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A>Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A:Reference number: A60143; MUID:85182650; PMID:2985580
 A:Accession: A60143
 A:Molecule type: protein
 A:Residues: 740-744,X,746-747 <RUS>
 R:Motczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A:Reference number: A38023; MUID:84191554; PMID:6325948
 A:Accession: A38023
 A:Contents: annotation; receptor activity
 A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Weisb, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A>Title: Functional independence of the epidermal growth factor receptor from a domain A
 A:Reference number: A33331; MUID:90003233; PMID:2790960
 A:Accession: A33331
 A:Contents: annotation; internalization signal
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C:Gene: GDB:EGFR
 A:Gene: GDB:EGFR
 A:Cross-references: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase; signal sequence; status predicted <SIG>
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1210/Product: EGF receptor #status predicted <EXT>
 F:25-65/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EB1>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EB2>
 F:446-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128,175,352,413,444,528,603/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:745/Active site: Lys #status experimental

Query Match 45.64; Score 3118; DB 1; Length 1210;
 Best local similarity 49.54; Pred. No. 3e-120;
 Matches 630; Conservative 178; Mismatches 349; Indels 116; Gaps 25;

```

11 LLLALLPPGA--STVCTGTMTKLRLPASPEHLDLRLYQCCQVYVGNLELYLPTN 68
14 LLAALLPPASRLAEKRYCCGCTSNKLTQLGTFEDHFLSLQRMFNNECEVALNLLEITYQRN 73
69 ASLSFLDIOEVGVYLIANNOVROVPLQRLIVRGQLFEDVYALAVLDNGPLNNTTP 128
74 YDLSFLKTIQEVGVYLIANNOVROVPLQRLIVRGQLFEDVYALAVLDNGPLNNTTP 126
129 VTGASPGGLRELOLRSLTEILKGVLIQRPOLCYOBTILMKDI FHKNNQALTLIDTNR 188
127 ---ANKTGKELPMRLQGLIHGAVRFSNNPACNVESTIOMRDISSDFLSNMSMDPQNH 183
189 SPACHPGSPMGKSCRCWGESSEDDQSLTRTVAGGCA-RCKGPLPTDCHEGCAAGTGP 247
184 LGSQCKDCSCPGSCWGAEEENCKLTKIICAQCSGRCGRGSPDCCHNCAAGCTGP 243
248 KHSDDLACLPFNHSGICELHCPALVYVNDTFESMNPBERRYFGASCVTACPYNYLSTD 307
244 RSSDCLVCKKFRDEATCKOTCPPLMLYNPTTYQMDVNPBEKYSFGATCVCKCPRNVVVD 303
308 VGSCTLVCPHNOVTAEDGTORCEKSCRCPCARVAVGLGMQYIKANSKFIQTELE-FAG 366
304 HSCSVACACADSVEM-EEDGVKRCCKCEGCRKVCNGIGIGERK-DLSLNATNIGHFKN 361
367 CKKIGSLAFLPESFGDPASNTAPLQPELOLVFLEETITGYLYISAMPDLSLDISVQ 426
362 CTSISGDLHLPLPAFGDSFTHTPPLDPOBLDLTKVKEITGFLLIQAMPENRTDLHAF 421
427 NFOVIRGRILHNGAVYLTQGLISWLGRLSELGSLALIHNNHLCVHTVPMDOLF 486

```

Db 422 NLEIRRTQKHQGFSLAVSLNITSLSGLSLKEISDGVIIISGNKMLCYANTIMKKLF 481
 Oy 487 RNNFTVSMFLRVKVSASHLE-ECVSEGLACHOLCARGHMGPGPTQCWCNSQFLRGE 545
 Db 482 GTSG-----QKTKIISNKGNSCKATGVCYCHALCSPEGCGEPDRDVCNRVSRRE 534
 Oy 546 CVEBCRYLQGLPREYVARNHLCPCHPREOPNGSVTCFGEPAADCVACAHYKDPFCVAR 605
 Db 535 CVDCSKLLEGPREFVENSECTIQHPRLPQAMNITTCGRPDNCTIOCAHIIIDGHCVTK 594
 Oy 606 CPSCVAKDLSYMPKRPDEGACOPCPINCTHSCVDLDDKCPAEOBASPLTISVAV 665
 Db 595 CPACVMEENNTL-VMKYADAGHYVCHLCPNCTYCTGPGLEGCEPTNGPKIP--STATGV 651
 Oy 666 G---ILVVVLGVVGLIKRQOKIKRYMRLLOETELVEPLTPSCAMPNQAOMRLK 722
 Db 652 GALLILVVALGIG--LFMRRIHVKRKLRLLOEELVEPLTPSGEAPNALRLK 708
 Oy 723 ETLRKVKVLSGAGFTVYKGIWIPDEENYKIPVAIKVLRNTSPKANKELIDEAYVAG 782
 Db 709 ETEFKKIKVLSGAGFTVYKGIWIPDEENYKIPVAIKVLRNTSPKANKELIDEAYVAG 768
 Oy 783 VGSFTYVRLIGICLTSTVOLVTOIMPYGCLLDHYRNRGRIGSODLLMCMQIAKMSYL 842
 Db 769 VDPHVCRLIGICLTSTVOLVTOIMPYGCLLDHYRNRGRIGSODLLMCMQIAKMSYL 828
 Oy 843 EDRLVHRDLAARVNLKSPNNHYKITDPEGLARLLDDETEYHAGGVVPIKMLLESTLR 902
 Db 829 EDRLVHRDLAARVNLKSPNNHYKITDPEGLARLLDDETEYHAGGVVPIKMLLESTLR 888
 Oy 903 RRFTHSDVMSYGVTEWELMTFGAKPYDGI PAREIPDLEKGERLPQRPCTIDVYIMV 962
 Db 889 RITTHSDVMSYGVTEWELMTFGAKPYDGI PAREIPDLEKGERLPQRPCTIDVYIMV 948
 Oy 963 KCMNIDSECPRRRELVESESRNARDQRFVVIG-NEDLSPASPLDSTFYKSLLEDDMG 1021
 Db 949 KCMNIDSECPRRRELVESESRNARDQRFVVIG-NEDLSPASPLDSTFYKSLLEDDMG 1008
 Oy 1022 DLVDAEYLYVPOGFCPPDRPAPGAGVNHHRHSSSTRSGGDLTLLEPSEEARPSPL 1081
 Db 1009 DLVDAEYLYVPOGFCPPDRPAPGAGVNHHRHSSSTRSGGDLTLLEPSEEARPSPL 1034
 Oy 1082 APSEGASDVFPDDDLGAKAGLQSLPTHPSPLOKRYSEDPVLPSET--DGVAAPLTC 1139
 Db 1035 LSSLATSN--NSTVACIDNGQSCPIKEDSFLQRISSPTGALTDSIDDTFL----- 1087
 Oy 1140 SPQPEYVNDVPRPQPSREGPLPAARPAATLERAKTSLPGKNGVYKOVFAFGAVEN 1199
 Db 1088 -PVPEYINQ-SVPKRPAGVQNPVYHNP--APSRLDPHYQD--PHSTAVGN 1135
 Oy 1200 PEYL-TPQGAAPQHPPRPAFSPAFDNLVYWDQ-----DP-----PERGAPST 1242
 Db 1136 PEYLNTVQ-----PTCVNSTFDSPAHMAKQSHQSLDNPYQODFEFPKAKENG 1186
 Oy 1243 FKGTPTAENPEYL 1255
 Db 1187 FKGS-TAENAEYL 1198

RESULT 5
 A53183
 Epidermal growth factor receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
 C:Accession: A53183; A43818; S24942; A28943; S45325; I49643
 R:Uetters, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.
 Genes Dev. 8, 399-413, 1994
 A>Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
 A:Reference number: A53183; MUID:94170986; PMID:8125255
 A:Accession: A53183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <LDE>
 A:Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A>Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A:Reference number: A43818; MUID:91232666; PMID:2030916
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Esinger, D.P.; Serrezo, G.
Submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A>Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hibbe, M.L.; Dunn, A.R.; Alexander, W.S.
Submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CA55587.1; PID:9488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A>Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: I49643; MUID:93126380; PMID:7678348
A:Accession: I49643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:I06864; NID:9193001; PIDN:AA53029.1; PID:9567201
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphoric
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
F:997, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.1%; Score 3086; DB 2; Length 1210;
Best Local Similarity 49.4%; Pred. No. 6e-119;
Matches 632; Conservative 170; Mismatches 358; Indels 120; Gaps 27;

11 LILALPFGAA--STVCTGTDMLRLPASPETHLMDLRLHYCCOVVQNLLETLYPTN 68
14 LILALCAAGALEBKVKCGQSTSNRLTOLGTFEDHFLSLQRMVNNCEVLANLETYYORN 73
69 ASLSFLQDIEQVQGVYLIANNOYRQVPLQRLRIYKGTQLFEDNYALAVLNDGPDPLNNTTP 128
74 YDLSPFKTIEQVAGVYLIANTVERIRPLEMLQIIRGNALYENTYALAILSN----- 124
129 VTGASGCGARELDRLSTETLKGVYLIQRPOLCYQDTILMKDI----FKKNQALATLI 184
125 -YGTNTRGRELRLPMRLQELILGAVRPSNPIICNMDTIQMRDIVQNVFWSNMSMD--- 180
185 DTRSRACHPCSPMKSGSRGSGSSSEDCSLTRTVACGGA--RCKGGLPTDCCHEGCAAG 243
181 -QSHPSRCRKCDSPCRGSGCWGGGEENQULTIKIICQOCSHRCKRGRSPSDCHQCNAG 239

244 CTGPKSDCLACLPFNHSGICELCPALVYNTDTFESMPNREGRYFGASCYTAAPVNY 303
240 CTGRRESDLVCQRFQDEATCTCPMLYNPPTTYQNDVPEGRYSGATCVKCKCPNRY 299
304 LSTVGSCTVCPPLHNOETVAEDGTORCEKSCPKCARCYGLGMQYIKANSKFIGITELE 363
300 VYTDHSGCVARCGDYIEV--EEDGIRCKKCKDDGCRVNCNIGIGERF-DLSTINATIK 357
364 -FACCKKIFGSLAFLEPSFGDPAASNTAPLOPBOLOVETLEBITGVYIYISAMPDSDL 422
358 HFKYCTAISGDLHLTPAFKGDSTFRTPPLDPRLEILTKYKEITGFFLLIQAMDNDWDL 417
423 SVFQNLQVINGRIILHNGASLTLOGIGISWLGRLSELGSLALIHNTHTLCFHTYPM 482
418 HAFENLTIIRKTKQHOQFSLAVVGLNITSIGLSLEISGDVYIISGNRLCYANTINW 477
483 DQLEFRPNFTVSWFLRVPKVSASHLE-ECVGEGLACHQLCARGHGCGPQCVNCSQFL 541
478 KKLFTGN-----QKTKIMNRAEKDCRVNHVCNPLCSSECGKPREPDCSCNVVS 530
542 RGQECVEECRYLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPF 601
531 RGRECVCKCNILEGEPRFVENSECICQHCPLQAMNITCTGRCGPONCICAHYIDGP 590
602 CVARCPGVPRDLSYMDIWKPRDEGACQPCPINCTHSQVLDKCKCPAEGRASPLTSIV 661
591 CVKTCPRAGINGENNTL--VMKTRADANNVCHLCHACTYGCAGPGLOGCEWPSGKPIBSIA 649
662 SAVGILLVVLVGVFGI-LIKRQOKIRKTYMRRLLOETELVEPLTPSGAMPQAOMRI 720
650 TGIYGGLLFIV--VALGIGLFMRRIIVKRTLRLLQEBELVEPLTPSGAPQAHRI 708
721 LKTELKRVKYLSSGAFGTYYKGIWIPDGENVKIPVAILKRENTSPKANEILDEAYVM 780
709 LKETEPEKKIKVLSGAFGTYYKGIWIPDGENVKIPVAILKRENTSPKANEILDEAYVM 768
781 AGVSPVSRLLGLCLSTVQVLTQMPYCGCLLDHVENGRGSOQLNNCMQIACGMS 840
769 ASVDNPRVCRLLGICLSTVQVLTQMPYCGCLLDHVENGRGSOQLNNCMQIACGMS 828
841 YLEDRVLRHRLAARNLVKSPNHVKITDFGLARLIDETEHYADGKVIKMALESI 900
829 YLEDRRLVHNDLAARNLVKTPQVKTITDGLAKLGAEEKHYAEGKVPKMALESI 888
901 LRRRTQSVMSYGVTVWELMTFGAKPYGCIIPAREIPDLLEKGERLPQPICTIDVMI 960
889 LRRITQSDVMSYGVTVWELMTFGSKPYGCIIPASDISILEKGERLPQPICTIDVMI 948
961 MVKCMIMDSRCRPRPRELVEEFSRMAPDPRFVYIC-NEDLGASPLDSTFFRSLLEDD 1019
949 MVKCMIMDADSRPRFRLELIEFSQMAPDPRVYLTQGDERRMHLPSPTDSNPFYRALMDEED 1008
1020 MGDVDAEYLVPOQGFPCPDPAAGAMVHHRRSSSTSGGDDLTLGLEPSEEARPS 1079
1009 MEVDVDAEYLVPOQGF-----NSST-----SRT 1034
1080 PLAPSBGASDVFDGLGMAKAGLSLPTHDSPLQRYSEDPVPLPSET--DGVAAPL 1137
1035 PLSTSLATSN---NSTVACINRNGSCRYKEDAFIORYSSDPGAVTEBNDIDAFI--- 1087
1138 TCSPOPEVYNQPRVQRPSPREGPLPAARAPADATLEBATLSPGKNGVYKDVFAFGAV 1197
1088 ---PVEEYVQ--SVKPRAGSVQNPVYHNPRLP-----APGRLLYON--PHSNAY 1133
1198 ENPEYL--TPGGAAPRPHPRPAPAFDNLVYMDQ-----DP-----PERGAP 1240
1134 GNEEYLVNTAQ-----PTCLSSGFNSPALMIQSGHMSLDNDYQDDFFPKETKPN 1184
1241 STFKGTPTANPEYLGIDVP 1260
1185 GIFKGG-PTAENAEYLRVAPR 1203

RESULT 6

TVCHLV

epiderma: growth factor receptor precursor - chicken

N:Contins: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C/Accession: A27720; A00643

R/Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart

Mol. Cell. Biol. 8, 1970-1978, 1988

A/Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A/Reference number: A27720; MUID:88261272; PMID:3260329

A/Accession: A27720

A/Molecule type: mRNA

A/Residues: 1-1223 <LAX>

A/Cross-references: GB:M20386

R/Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rotman, F.M.; Crittenden, L.B.; Raines, M

Cell 41, 719-726, 1995

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Accession: A00643

A/Molecule type: mRNA

A/Residues: 585-1223 <NLL>

A/Cross-references: GB:M10066

C/Genetics:

A/Gene: erbB

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor

specific protein kinase

F/1-30/Domain: signal sequence #status predicted <SIG>

F/31-1223/Product: epidermal growth factor receptor #status predicted <MAT>

F/31-654/Domain: extracellular #status predicted <EXT>

F/81-307/Domain: EGF receptor extracellular domain repeat <EE1>

F/197-610/Domain: EGF receptor extracellular domain repeat <EE2>

F/655-677/Domain: transmembrane #status predicted <TM>

F/678-1223/Domain: intracellular #status predicted <INT>

F/719-984/Domain: protein kinase homology <KIN>

F/727-735/Region: protein kinase ATP-binding motif

F/136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #

F/192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted

F/687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F/754/Active site: Lys #status predicted

F/1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.9%; Score 3070.5; DB 1; Length 1223;

Best Local Similarity 48.2%; Pred. No. 2.6e-118;

Matches 629; Conservative 175; Mismatches 347; Indels 153; Gaps 28;

QY 8 RMGLLALPFGAA-----STVCTGTMKRLRLPASPTHLDMLRHLVGGCQVVGNGLE 61
 DB 13 RGAAILVLLGLVALCSAVEEKVCOGTNNKLTQLGHVEDHFTSLQRMNNCEVJLSNLE 72
 QY 62 LTVLPFNASLSFLODIOEVGVYLIANOVROYPLQRLTVRGQLPEDVYALAVLNGD 121
 DB 73 IIVENHRDITFTKIOEVAGVYLIANMVDVIPLENLQIRGVNLVDSFALAVSNYH 132
 QY 122 PLNNTPVTGASPGRLQRLSLTELKGVLIQRNPOLCYODTILMKDIFPKNNQAL 181
 DB 133 -NMKTQ-----GRLRLPKRLSELINGVKSINPKLCMDTYVAMNDIITSRK-PL 182
 QY 182 TLID-TNRBACHPCSPMGKGRGWSSESDCSLRTVCAGCA-RCKGPLPTDCHEQ 239
 DB 183 TVLDFASNLSSCKPCHNCTEDHCWGAGEONCQTLTVKICAOQCSGRGKVPSSDCCHN 242
 QY 240 CAAGCTGPHGSDCLACHFHSGICELHCPALVTYNTDFESMPNPRGRTFGASCYTAC 299
 DB 243 CAAGCTGPHGSDCLACHFHSGICELHCPALVTYNTDFESMPNPRGRTFGASCYTAC 302
 QY 300 PNYLSTVDSGCTLVCPRLHNOETVADGTQRCCKSKPCARVCYGLGMQYIKANSKFIGI 359
 DB 303 PHNYVTVTDHSGCYRSCNTDTYEV-EENGVRKCKKCGGLSKVNGICIGLKLILS-INA 360
 QY 360 TELF-PAGCKKIFGSLAFPLPESFDGPASNTAPLOEBOLOVFTLEITGYLYISAMPDS 418

DB 361 TNIDSFKNCTKINDVSLIPVAFGLDAFTKTLPLDPKKLDVFRVKEISGELLQAMPDN 420
 QY 419 LPDLSVONLOVIRGRILHNGAYSLTQGLGISWGLRSLEISGLALIHNNTHLCVH 478
 DB 421 ATDLYAFENLEITGRKQKHQOISLAVNLTQISGLSLSEISDGIAMKNNKLYCAD 480
 QY 479 TVPMDQLFRFNNPVSFMLRVPKVSASHLECEVGLACHQLCARGHGCGPTQCVNCS 538
 DB 481 TNNRSLFATQS-----QTKIIONNNKNDCTADRHVCBDPLCGDVGCMGCGPFHCFSCR 534
 QY 539 QFLRGQCEVCEVYLOGLPREYVNRHCLPCHPECPQNG--SVTCFGEPAQCVACAH 595
 DB 535 FFSRQKCEVQCNLTQGEPRFEFDSKCLPCHSECLVONSTAYNTTCSGPPDHCMKCAH 594
 QY 596 YKDPFPVACCPSSVKNRDLSTMPYKRPDEGAQOQPCPINTHSCVLDLDDGCAEQAS 655
 DB 595 FIDPFPVACCPAVLGEENDTL-VWKADANAVQQLCHPNTCRCKGGLGCGP---NGS 650
 QY 656 PLTSIVAVV-GILLVVVLGVVFGILIKRQOKIRKTYMRLLOETLVEPLTPSGAMPN 714
 DB 651 KTSIAGVGVGLCLVVGIGLGLYLRH-HYRKRTLRLLDERLVEPLTPSGAMPN 709
 QY 715 QAKRILKTELKRVKVLGSGAFGVYKGIWIPGENVKIPVAIKVRENTSPRANKEL 774
 DB 710 QAHRLIKETEFKKVYLGSGAFGVYKGIWIPGENVKIPVAIKVRENTSPRANKEL 769
 QY 775 DEAVTMAGVSPYSLRILGTLSTVQLVQMLMYGCLLDHVRNRRGLSGODLNMCMQ 834
 DB 770 DEAVTMAGVSPYSLRILGTLSTVQLVQMLMYGCLLDHVRNRRGLSGODLNMCMQ 829
 QY 835 IAKMSYLEDVRLVHRLAARNVYVKSPPNHYKIDFGARLDIDETEVHADGGVPRPKM 894
 DB 830 IAKMNTLEERLRLVHRLAARNVYVKSPPNHYKIDFGARLDIDETEVHADGGVPRPKM 889
 QY 895 MALESILRRPFTHQSDVMSYGVTVWELMTFCAKPYDIPAREIDLEKGRLEPPRPT 954
 DB 890 MALESILRRPFTHQSDVMSYGVTVWELMTFCAKPYDIPAREIDLEKGRLEPPRPT 949
 QY 955 IDVYMTWKCMVIDSEKRPFRFVLSFSPKMARPPQPFVYIQ-NEDJGAPSLDSTYRS 1013
 DB 950 IDVYMTWKCMVIDSEKRPFRFVLSFSPKMARPPQPFVYIQ-NEDJGAPSLDSTYRS 1009
 QY 1014 LLEDMDGDLVDAEYLVVPOQGFPCPDPAQAGAGMHHRRSSSTRGGDLTLGLESE 1073
 DB 1010 LLEDMDGDLVDAEYLVVPOQGFPCPDPAQAGAGMHHRRSSSTRGGDLTLGLESE 1038
 QY 1074 EEARPSPL-----APSEGASDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVPLPS 1128
 DB 1039 ---SRTPLLSSLSATSNNSATNCID-----RNGGHHVREDSFYQRYSSDPTGNFLE 1087
 QY 1129 ET--DGVVAPLTCSPQPEYVNOQDVPQPPSPREGLPAPARAGATLERAKTISPQNGV 1186
 DB 1088 ESIDIDGFL-----PAPEYVNO--LMPKKS-----TAMVQNOI 1118
 QY 1187 VKDVF-----AFGAVENPEYLTPOGGAAPHPHAFSAFENDLYWQD 1231
 DB 1119 YNNISLTAISKLPMSDRYONSHSTAVDNPEYL-----NTQGPSPLAKTVFESSPYWQ 1170
 QY 1232 -----DPEE-----RGAPPSTFKGPTAENPEYLGIDVY 1260
 DB 1171 SGNHQINLNDPQVQDFPLPHTKPNGLKVPAAENPEYLVAAAP 1214

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C/Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C/Accession: A47253

R/Flouman, G.D.; Cuiousecou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A/Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderma

A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <P/O>
A:Cross-references: GB:L07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A:Note: sequence extracted from NCBI backbone (NCBI:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.0%; Score 2942.5; DB 2; Length 1308;
Best Local Similarity 45.0%; Pred. No. 4.7e-113;
Matches 609; Conservative 181; Mismatches 383; Indels 179; Gaps 29;

9 WGLLLALLPGAA---STOVCTGTDMKRLRPASBETHLMDLRHLYGCGVAVQGNLELTLY 64
8 WVVVSLVLAAGTVGPDSQSVACGTENKLSLSLDLQGYRALRKYYENCEVVMGNLEITS 67
65 LPTNASSFLDIOEVGVYLIANHNOVPLQRLRYRGTOLEPDNVALAVLDGDEPLN 124
68 IEHRDLSFLRSVRELVGVYVALNQFRYLLPLENRIIRGTLYEDRYALAIPLNVRKDG 127
125 NTFVTGASPGGLRELQRLSTELIKGGVLLIQNPOLCYODTIIMKDIYHKNQALATLI 184
128 NF-----GLOELGKMLTEILNGGVYVDQNFELCYADTIHMDIIVRMPENLTLV 178
185 DTNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVACGC-ARCKGPLPTDCHEQCAAG 243
179 STNGSSGCGRCHKCTG-RCWGPTEHNCQTLRTVCAECQDRCYGPVSDCHRECAAG 237
244 CTGKHSPCLACLPNHSIGCEIHCPALVTYNTDFEEMPNDEGYTGAQVTAQPNY 303
238 CSGPRLDTCFACMNFNDGACTQCQTFVYNPTTFQLEHFNMAKYTGAFVKKCPNPF 297
304 LSTVGSCTLVCPPLHNOEVTADGTQRCCKSKPCARCYGLGMQYIKANSKFIGITILE 363
298 V-VSSSCVTRACPSSKMEV-BENGIKMKCPCTDICPKACDGICTGSLMSAQVNVSSNDK 355
364 FAGCKTIFGSLAFLEPESFDGPASTARLAPOLQVFTLEBTGYLYIASMPDLPLS 423
356 FINTCKINGNLIPLVATGJHGPYNAIEAIDEXKLVNFTVEITFPLIQSGPMPMTDPS 415
424 VFQMLQVIRGRILHNHGAVALTLOGIGISMLGRLSRLREISGALILHNHTHLCFVYTVMD 483
416 VFSNLVITIGRVLISGLSLTLIKQGITSLQFQSLKETSAGITVITDSSNLCTYHTIMWT 475
484 QLFRFNNFTVFWLVRPVKVSASHLEECVGEGLACHQLCARGHCWGPCTQCVCNCSQPLRG 543
476 TLFSTIN-----QRIVIRDNRKANNCNCTAEGMVCNHLSSQGCWGPDPQCLSCRFRSG 529
544 QECVBECAVLOGLPREYVNAHCLPCHRECP-ONGSVTCTGPREADQVCAAHYKDPFC 602
530 RICIESCNLYGGEFEFENGSIQVECDPCCEMDEGLTTCGPGPDNCTCKSHFKDGPNC 589
603 VARPSPGVKPDLSYMPRIKPPDEBAGQPCPINCCHSCVDDDKSC-----PAE 651
590 VEKPPDGLQGANSF--IFKYADPDECHPCHPNCTGCGNCPGSHDCTIYPTGHSITLPOH 647
652 QRASPLTSIVAVV-GILLVVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTSPG 710
648 AR-TPL--IAAGVIGGLFIVLIVGLTFVAVVRRKIK-KGALARFL-ETELVEPLTSPG 702
711 AMPNQAQRIILKETELRKVKVLGSAFGTYVKGIMIPGENVKIPVAIKVLRENTSPPAN 770
703 TAPNQAQRIILKETELRKVKVLGSAFGTYVKGIMIPGENVKIPVAIKVLRENTSPPAN 762
771 KEILDEAVYMAVGSPYVRLIGLICTSTVQLVTLQMLPEYGCILDHVRNQRGLSGDILN 830
763 VEPFDEALIMASMDHPHLVRLILGVCLSPTIQLVTLQMLHGCILEVYVHKKONISQILLN 822
831 WCMQIANKMSYLEDVRLVHRDLAARNVLKSPNHVKITDPGLARLLIDETETYNADGGKV 890

DB 823 WCMQIANKMSYLEDVRLVHRDLAARNVLKSPNHVKITDPGLARLLIDETETYNADGGKV 882
DB 891 PIKMALESILRRRFTHSDVWSYGVYWEIMLTGAKYDGIIPAREIDLLEKGRLLPQ 950
DB 883 PIKMALECIHYRRFTHSDVWSYGVYWEIMLTGAKYDGIIPAREIDLLEKGRLLPQ 942
DB 951 PICTIDVYMIWKCMMIDSECRPRPRELVSEFSWAPRQRFVYVONED-LGPA5PLUST 1009
DB 943 PICTIDVYMWKCMCIDADSRPKEKLALESFRRARDPQRYLVIGQDRMLPSPNDK 1002
DB 1010 FYRSLLEDDMDGLDVAEEYLVPOQGFPCPPAPGAGGMVHRRSSSTRSGCDLTGL 1069
DB 1003 FFGVNLDEEDLEDMDVAEEYLV-QAFNIPPP-----IYRSRAIDNRS-----ELGH 1050
DB 1070 EPSEBEAPRS-----PLAP-SEGASDVFDGLGMA 1100
DB 1051 SPPAYLTWMSGNQFVYRDGFAAGCVSVYPAPRTSTIPEAPVAGATAEILFDDSCNGT 1110
DB 1101 AKGLQSLPTHDPSPLOYSDEPTVPLPS-----ETDGYVALPLCSPOPEVYNQPDVAP 1153
DB 1111 LRKVPAPHVQDSSSTORYSAOPTVAPAPERSRGELDEGGYMTPMKDKKQELNVE-- 1167
DB 1154 QPSPREGRPLPAPAGATLEBAKTLSPKKGWVVDVAFGGAIVENPEYLTFGCGAARQ 1213
DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNPETHNASNG----- 1194
DB 1214 HPPPA-----FSPAFDNLVYWDPPERGA- 1238
DB 1195 -PPKAEDEYVNEPLYLNTFANTLGAKEYLKNILSMPEKAKKAFDPNWMHSLPPRSTL 1253
DB 1239 -PPSTFKCTPT-----AENPEYL 1255
DB 1254 QHPDYLOEYSTKYFYKNGRIPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
N:Alternate names: epidermal growth factor receptor homology; kinase-related transformi
C:Species: Xiphophorus maculatus (southern platyfish)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
R:Witbrodt, J.; Adam, D.; Maltischek, B.; Meunier, W.; Raulf, F.; Telling, A.; Rober
Nature 341, 415-421, 1989
A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu 1
A:Reference number: S06142; MUID:90015140; PMID:2797166
A:Accession: S06142
A:Molecule type: DNA
A:Residues: 1-1166 <WT>
A:Cross-references: EMBL:X16891, NID:G65290; PIDN:CAA34770.1; PID:G65291
Oncogene 6, 73-80, 1991
R:Adam, D.; Meunier, W.; Scharf, M.
A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph
A:Reference number: S13807; MUID:91125882; PMID:1846957
A:Accession: S13809
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A:Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA3763.1; PID:G65285
C:Genetics:
A:Gene: mrk
A:Map position: Y
A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; ty
F:1-25/Domain: signal sequence #status predicted <SIG>
F:126-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 38.9%; Score 2661; DB 1; Length 1166;
Best Local Similarity 44.8%; Pred. No. 1.3e-101;


```

Db      238 PODTDFACRHFNDSGACVRCPOPLVYNKLTFOLEBNPHTKYQYGGVCAASCPHNFV-V 296
Qy      307 DVGSCITLVCPILHNOEYVAEDTORCEKCSKPCARVCGIGLMQYIKANSKF--IGITBLE- 363
Db      297 DQTS CVRACPPDKMEVD--KNGLKMCBPCGGICCPACCEGTG----SSRQYVDSNIDG 350
Qy      364 FACCCKIFGSLAF.PESFPDDPASNTAPLOPEOLQVETLEETLEITGYUISAMPDLS 423
Db      351 FVACTKILGNLDFLITGLNDPMMHKIPALDPEKLVNFRITREITGYINIGSWPRPHMFS 410
Qy      424 VFONLOVIRGRILHNGAYS--LTLQGLGISWLGRLSRLSGSLALIHNTLHLCFVHTVPW 482
Db      411 VFSNLTITIGRSILYNRGFSLLIMKNLNVTSIGFRLSLEISAGRIYISANQOLCYHSHLN 470
Qy      483 DOLFERNFVFWLRYP---KVSASH---LEGVNGSLACHQICARHGCMGPRPTCVN 536
Db      471 TKY-----LRGPTEERLDIKNRRPRDDVAAGKVCDFLCSSGGCMGPRGQCLS 519
Qy      537 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEPAEQVACAHY 596
Db      520 CRNYSRGVCVTNCHNPLNGEPREPAHEACFSCHPEQPMEGTATCMGSGSDTCAQCANH 579
Qy      597 KDRPFCAACPSGVKPDLSYMPIMKPFDEBACOPCPINCTHSCVDLDDKGCRAQRA-- 654
Db      580 RDPHCVSSCPHGVLG--AKGPIYKYPDVONECRPCHENCTOGCKGPELDCLQOTLVL 637
Qy      655 --SPLTISAVVGIILVVVLGVFGLIKRROOKIR--KYTMRLLEOTELVEPLTSGA 711
Db      638 GKTHLTALTAVING--LVVIFMLGTFGLWGRRIQNKAMRYLRGESIEPLDS-E 694
Qy      712 MPQAQMRILKETELRKVVLGSGAFGTYYKGIWIPGENVYKIPVAILKVLRENTSPANK 771
Db      695 KAKKVLARIKETELRLKVLGSGVFTGHKGVMIPEGESIKIPVCIKVEDSGRSQSF 754
Qy      772 EILDEAVWAGVSPPYSLIGLICTSTVOLVQLMYCGCLLDHVENKRGSLSDILLN 831
Db      755 AVTDHMLAIGSLDHAHIVRLGLCPGSSQLVTOYELPGLSLDHWORHGLGQLLLNW 814
Qy      832 CMOIAKMSYLEEDVRLVHRDLAARNVLSKSPNHKIIDPFGIARLDIDETEHADGKVP 891
Db      815 GVQIAKMTYLEEHGVAHRLAARNVLLSKSPSOVQVADLPFGADLLPPDDQQLYSEAKTP 874
Qy      892 IKKMALESILRRBFTHQSDVMSYGVTVWELMTFGAKPYDGI PAEPIPDLEKEGRLLPQP 951
Db      875 IKKMALESIFHGKYTHQSDVMSYGVTVWELMTFGAEFYAGIRLAEVOLLEKEGRLLAQ 934
Qy      952 ICTIDVYIMVWKCMVIDSECRPRRELVSFSPMAARDPQRFVITQNEIDLPA--SPLDS 1008
Db      935 ICTIDVYMWVWKCMVIDENIRPTEKELANEFTRMARDPRLYLVIKRES--GFGIAPREPH 993
Qy      1009 TFRSLLEDDMDGLVDABEYLVPOQGFCCRDPARAGAGVNHHRHSSSTRSGGDLTLG 1068
Db      994 GLTNKLEVELEBEDLDLDLEAEED-----NLATTTIGSALSILP 1034
Qy      1069 LEP--SEEDARSPPLABEGAGSDVFDGLCMGAKGLQSLPTHD--PSPLORYSEDPVPL 1126
Db      1035 VGTILNPRGOSILSPSSGY--MPMNQNLGESCOESAVSSSRSCRPRVSLH-----PM 1087
Qy      1127 P-----SETDGYVA-----PLTSGQPR-----YNOQPRVRRPOPSPRE 1160
Db      1088 PRGCLASESEGHVTSSEALQEKVSMCRSRSRSPRPGSDAUYHSQRLSLTPVPLS 1147
Qy      1161 GP-----LPAARPAGATLEBRAKTLP--GKNGV-----KQVFAFGAVNREY 1202
Db      1148 PROLEBEDVNGYMPDTHLKGTSSREGTILSSVGLSVLTBEEDD-----EYEV 1199
Qy      1203 LTPQGAAPQHPAPPAFPADNLVYWD-----QDPPERCASTPTKGTPTAA 1250
Db      1200 MNRRRHSP--PHPPRSSLEELGYEYWDVGSDLASLIGTQSCPLHPVPIPTAGTTPDE 1258
Qy      1251 NPEYL 1255
Db      1259 DVEYIM 1263

```

```

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N/Alternate names: ErbB3 protein; HER3 protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C/Accession: J04387
R/Heliyer, N.O.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A/Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein
A/Reference number: J04387; MUID:96096535; PMID:8522190
A/Accession: J04387
A/Molecule type: mRNA
A/Residues: 1-1339 <HEL>
A/Cross-references: GB:U29339; NID:9915389; PID:9915390
A/Experimental source: liver
A/Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue
C/Comment: This protein is a functional heregulin receptor that transduces signals to
C/GeneID: 8163
A/Gene: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homo
C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #stratus predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #stratus predicted <MAT>
F:640-659/Domain: transmembrane #stratus predicted <TM>
F:705-970/Domain: protein kinase homolog <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)
Query Match 33.64; Score 2300.5; DB 2; Length 1339;
Best Local Similarity 40.18; Pred. No. 7.5e-87;
Matches 520; Conservative 171; Mismatches 430; Indels 175; Gaps 36;
Qy      3 LAAACRHWGLLALIPGAA---STQVCTGDMKRLPASPETHLDMLRHLYOGGVQGN 59
Db      7 LQVLC-----FLSLARSEBNGNSQAVCPGLNLSTVGDDANQVQTLKYKEKEVWGN 62
Qy      60 LELTLEPTNASSFLQDIOEVQGVLIANHNVQVPLQRLRIYRGTLFEDNVALAVLDN 119
Db      63 LEIVLTGHNADLSFLQWIREVTAVLVANKEFSLPLNLRVVRGTVQVDOCKFAIFVM-- 120
Qy      120 GDLNLTPTVYTGASPGGLRLQLRLSLEILLKGVTLIRNQLCYQDITLTKDIFHNKQL 179
Db      121 ---LNVNT---NSSHALRLQLKFTQLEILLISGVYLEKNDKLCMDITIDRRDIYRV-- 170
Qy      180 ALTLIDTNRGRACHPCSPMKSGRCWGESSEDCSLTRTYCAGGC-ARCKGRLPTDCHE 238
Db      171 GAEIVKKNANCPCHIEVKG--RCWGPGRDDCOILTKTICAPQCNRCGFPNQCCHD 229
Qy      239 QCAAGCTGPHGSDCLACLHNHSGICEHLCPALVTYNTDFESMPNDEGRYTFGASCUTA 298
Db      230 ECAGGCSGPDQDTCFCACRRFNDGACVRCPEPLVYNKLTFOLEBNPHTKYQYGGVCAAS 289
Qy      299 CPYNYLSTDVSGCTLVCPILHNOEYVAEDTORCEKCSKPCARVCGIGLMQYIKANSKF 356
Db      290 CPNHFV-VDTQFCVRCPPDKMEVD--KNGLKMCBPCGGICCPACCEGTGSSRQYVDSN 347
Qy      357 IGITLEFAGCKKIFGSLAF.PESFPDDPASNTAPLOPEOLQVETLEETLEITGYUISAMP 416
Db      348 ID-----GFVACTKILGNLDFLITGLNDPMMHKIPALDPEKLVNFRITREITGYINIGSWP 403
Qy      417 DSLPLDSVFONLOVIRGRILHNGAYS--LTLQGLGISWLGRLSRLSGSLALIHNTLHLC 475
Db      404 PHMNFVSFNSNLTITIGRSILYNRGFSLLIMKNLNVTSIGFRLSLEISAGRIYISANQOLC 463
Qy      476 FVHTVWDQLFRNNMTVSWFLRVPKVSASHLE-----ECVBEGLACHQLCARGRHGW 529
Db      464 YHSHLWVTRL-----LRGPSEERLDIYDRPBLGCLAEGRVCPDLCSSGGCMG 512
Qy      530 GPTQVNCGQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEPAEQ 589

```

Db 513 APQCGLSCRNYSRGVCTHGNPLQGBREPREFHNAQCFSCHPECLMEGISTYNGSSDA 572

Qy 590 CACAAHYKDPFVCYACRSPGVKPDLSYMPIMKFRPEEGACQPCPINCTHSC--VDLDKG 647

Db 573 CACAAHFRRDPRPHCVNSCPHGILG--AKGPYIKYPRDAQNECRPHENCTQCGNPELIDCL 630

Qy 648 CPAEQASPLSTISAVVGLLVVLGVVFGILLIRROOKIR-KYIMRRLQETELVEPL 706

Db 631 GQAEVLMSKPHLVIAVTVG--LAVLIMLQGSFLVWGRRIQNKRMRRYLERGESTEPL 688

Qy 707 TPFGAMPNOAQMRILKETELRKVKVLSGASFGSTVYKGIIPDGENVKIPAIKVLRENTS 766

Db 689 DPS-EKANKVLARIKTELRLKLVKSGVFTYHKIGIMIPEESIKIPICIVIEKSG 747

Qy 767 PKANKETLDEAYVAVAGVSPVSRLLGLCTASTVOLVTQIMPYGCLLDHYENRGRLSGQ 826

Db 748 RSQFOAVTDHMLVAGSLDHAHIVRLGLCPGSSLLDLVTQYLPGLSLDDHVKQRETLGPQ 807

Qy 827 DLLNMCQVAKGMSYLEDVRLVHRDLAARNVLVKSNNHKITDFGLARLLIDETETHAD 886

Db 808 LLLNMGQVAKGMVYIEESHVNRDLARNWLKSPSQVQVADFVADLLPPDDKOLLHS 867

Qy 887 GGVPIPKMALESILRRRFTQSDVMSYGVTWELMTFGAKPYDGIPIREIPDLEKGR 946

Db 868 EAKTPIKMALESILHFGKTYTHQSDVMSYGVTWELMTFGAEPYAGIRLAETIPDLEKGR 927

Qy 947 LPPPICTIDVYIMVVKCMWIDSECRPPRELVSFSRNARDPQRFVIONEDLGPASPL 1006

Db 928 LAQPOICTIDVYVWVVKCMWIDENIRPTFKELANFTRMARDPPRLVLIKRA-SGPQT- 985

Qy 1007 DSFTFYSLEDDMGDLVDAAEYLVPOQGFCDPAFGAGVNHHRSSSTRSGGDLT 1066

Db 986 -PAEBSVLTTKEL-----QEAELEPEL-----DLT 1010

Qy 1067 LGLEPSE-----EAPRPLAPSEG-----AGSVFPDDGLM 1098

Db 1011 LDLEAEBEGLATSLGASLSLPTGILTRPRGSSLSLSPSGYMPNMOSSLGEACLDLSAVLG 1070

Qy 1099 GAAGKQLSLPTHPDSPLORYSEDPVPLPSETDGYV---APV-----TC-----SPQ 1142

Db 1071 GREQFRRPILSH-PIPRGR-----PASESEGHVYTGSEAELOEKVSYCRSRSRSPR 1122

Qy 1143 PE---YVNPDPVARPPQSPPEGP-----LPAARPAQATLERAKTISP-GKNGV 1186

Db 1123 PRGDSAYHSQRHSLLTFTVPLSPSPGLEEDENGVMYMPDTHRLGASSRSRGTSLSSVGLSSV 1182

Qy 1187 V-----KQVFAFGCAVENPEYLTPOGGAAPQPHRP 1216

Db 1183 LGTEDEED-----EYEVYMNKRGRSP-PRPP 1209

RESULT 11

T/VF/V

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C:Species: avian leukosis virus, ALV

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999

C:Accession: B00643

R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.

Cell 41, 719-726, 1985

A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and protein

A:Reference number: A00643; MUID:85228222; PMID:2988784

A:Accession: B00643

A:Molecule type: mRNA

A:Residues: 1-698 <N/L>

A:Cross-References: GB:M10066, GB:M13881, NID:g211749, PIDN:AAA48763.1, PID:g211750

A:Note: in Genbank entry CKEKRB8F, release 109.0, the source is designated as Gallus gallus

C:Comment: This protein is synthesized as a gag-env-erbB protein.

C:Genetics:

A:Gene: gag-env-erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein

F:1-6/Product: gag protein (fragment) #status: predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-69/Product: protein-tyrosine kinase e2b8 #status predicted <ERB>
F:144-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match	25.8%	Score	1766.5	DB	1	Length	698
Best Local Similarity	52.2%	Pred. No.	2.5e-65				
Matches	374	Conservative	80	Mismatches	137	Indels	125
						Gaps	18

QY 584 GPEADQVCACAAHYKXOPPFVACRPSGVKRDLSMPILMFDEGACQCPCLNCHSICYDL 643

Db 60 GP--DHCKKCAHFIDGPHCVKACRPAVLGEQNTL-VKRYADANAACOLCHPNCITRGCKGP 116

QY 644 DDKGCPAEQASPLTSIVSAVV-GILLVVLGVVFGILIKRROOKIRKYMRLLOETEL 702

Db 117 GLEGCP--NSGKTPSIAGAVGVGILLCLVVVGIGIGLYLRR-HIVKRLRLRLQEREL 172

QY 703 VEPILTPSGAMPQAOQMRILIKETELRKVKVLGSGAFGVYKGIWPDGENVKIPVAIKYLR 762

Db 173 VEPILTPSGEAPQOAHRIKLETEPFKKVKVLGSGAFGVYKGLWIPBEKVKIPVAIKELR 232

QY 763 ENTSPKAKEILDEAYVYVAGVGSPPVYSRLIGTCLTSTVOLVTOQLMPGCLLDHRENGR 822

Db 233 EATSPKAKEILDEAYVYVAGVSDNPHVCRLLIGTCLTSTVOLITQIMPGCLLDYIREHKN 292

QY 823 LGSODLLMNCQOIAKMSYLEDVRLVHRDLAARNVYKSPNNHYKITDFGLARLLDIDETE 882

Db 293 IGSQVLLMNCQOIAKMSYLEERLVLHRDLAARNVYKTPONHYKITDFGLAKLGADEKE 352

QY 883 YHADGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLE 942

Db 353 YHAEQKVPKIMMALESILHRITHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSYLE 412

QY 943 KGERLPORPCTIDYVYIMVKCMATIDSECRPRRELVSERSMARDPQRFVVIQ-NEDLG 1001

Db 413 KGERLPORPCTIDYVYIMVKCMATIDSECRPRRELIAEFSKMARDPRLVLIQGDERRH 472

QY 1002 PASPLDSFFVYSYLEDDDDGDLVDABEYLVPOGFCPCRPAPACGGMVHNRHSSSTRSG 1061

Db 473 LPSPLDSFFYRTLMKEEDMEDIVDAEYLVPHGGF-----NSPST--- 513

QY 1062 GGDITLTGLPSEBEAPRSPL-----APSEGAGSDVDPDGLGMAKGLQSLPTHDSPSLQ 1116

Db 514 -----SRTPLLSSLSATSNNSATNCID-----RKGQGHPRVAEDSFVQ 550

QY 1117 RYSEDPTVPLPSET--DGVADLTCSPPQRYVYVQNPVRPQSPSPREGRLPAARPAAGATLE 1174

Db 551 RYSDPTGNFPLEESIDDGFL-----PABEYVYVQ--LMPKKPS----- 585

QY 1175 RAKTLSPKNGVVKQV-----AFGAVNEPEYVLTPOGGAAPQHPHPAPAF 1219

Db 586 ---TAAVQONQIYNNISLTATSKLPMDSRYQNSHSTAVNPEYL-----NTNQSPLA 633

QY 1220 SPAPDNLVYMDQ-----DPPE-----RGAPSTFGKPTLAENPEYLGIDLP 1260

Db 634 KTVESSESYVYIQSGNHQINLNDPDIQODPLFNETKPNGLKVPALAEPEYLRVAP 689

RESULT 12

TVYUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)

C:Species: avian erythroblastosis virus

C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999

C:Accession: A00644; A38022

R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.

Cell 35, 71-78, 1983

A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fam

A:Reference number: A00644; MUID:84026539; PMID:6313229

A:Accession: A00644

A:Molecule type: DNA

A:Residues: 1-604 <YAM>

A:Cross-references: GR:A01216; NID:G209676; PTDN:AAA32400 1. PTD:G209678

R, Debuitte, B.; Henry, C.; Benaissa, M.; Bisette, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984
A:Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28, 'F', '141-145', 'V', '147-152' <DB>
A:Cross-references: GB:K02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.5e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
QY 593 CAHYKDPFCVACRPSGVKPLSYMPIMKFPDEBACOPCPINCTHSCVDLDDKCGPAEQ 652
DB 3 CAHFDGPHCVKACPCAGVLGENDTL-VKRYADANVCQLCHPNCIRGCKGPGLEGC--- 58
QY 653 RASPLTSIVSAV-V-GILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGA 711
DB 59 NGSKTPSIAGVGVGLCLVVGVLGIGLYLRR-HIVKRTLRLLQREIIVEPLTPSGE 117
QY 712 MPNOAMRILKETELRKVKVLSGAFGVYKGIWIPDEENKIPVAIVAEINSPKANK 771
DB 118 APNOAHILKETEFKVKVLSGAFGTYKGLMPBGEKVIPIALILEBASPANK 177
QY 772 EILDEAVVAGVSGPYVSRLLGICLTSTVOLVQLMPYGCCLLDHVENRGLSGODLNM 831
DB 178 EILDEAVVAGVSDPNHVCRLIGICLTSTVOLITQMPYGCCLLDYREHKDNGSYLLNM 237
QY 832 CMQIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITPFGILRLIDETETHADGKVP 891
DB 238 CVOIAKGMNLYEERLVRDLAARVLYKTPQHVKITDFGLAKLGADKEKYHAEKGKVP 297
QY 892 IKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGI-PAREIPDLLEKGERLP 951
DB 298 IKMALESILRRITTHOSDWSYGVTVWELMTFGSKPYDGI-PASEISSVLEKGERLP 357
QY 952 ICTIDVYIMVCKMMDISECRPRRELVSERPMARDPQRFVIO-NEDLPASPLDSTF 1010
DB 358 ICTIDVYIMVCKMMDIDSRPKFRELIAPSKMARDPRLVIOGDERMHLPEPTDSKF 417
QY 1011 YRSLLEDDMDGLVDAERYLVPOGFCFDPAPAGAGMTHHRSSSTRSGGDLTGLE 1070
DB 418 YRSLMEEDMEDIVDADEYLVPHOGFF-----NSPST----- 449
QY 1071 PSEEARSPPL-----APSEGASDVFPDGLQMGAKGLQSLPTHPDRLQRYSEDPVP 1125
DB 450 -----SRPLLSLSLATSNNASATNCID-----RKGQGHVPREDSFVORYSSDPGN 495
QY 1126 LPSET-DGYVAPLTCSPQPEYVNOQPVRRPQSPREGPLPAPAPAGATLERRAKTLSPCK 1183
DB 496 FLEESIDGFL-----PAPRYVNO--LMPKKPSTAM----- 524
QY 1184 NGVVKDVFAP-----GAVENPEVLTQCGAALPQPPAPSPAFD 1224
DB 525 --VONOYVNFISLTAISKLPMSDRYONSHSTAVDNPEYL-----NTNOSPPLAKTVFE 574
QY 1225 NLVYWDODPPERCAPSTFKCTPAENPEY 1254
DB 575 SSPYWIQSGNHQ-----INLDNPDY 594

RESULT 13
S35745
C:Species: avian erythroblastosis virus
C:Species: avian erythroblastosis virus

C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
Submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.5e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;
QY 584 GPEADOCVCAHYKDPFCVACRPSGVKPLSYMPIMKFPDEBACOPCPINCTHSCVDL 643
DB 1 GP--DHCKCAHFDGPHCVKACPCAGVLGENDTL-VKRYADANVCQLCHPNCIRGCKGP 57
QY 644 DDKCPAQRASPLTSIVSAV-V-GILLVVVLGVVFGILIKRQOKIRKYMRLLOETEL 702
DB 58 GLEGC---NSKTPSIAGVGVGLCLVVGVLGIGLYLRR-HIVKRTLRLLQREI 113
QY 703 VEPLTPSGAMPOAMRILKETELRKVKVLSGAFGVYKGIWIPDEENKIPVAIKYLR 762
DB 114 VEPLTPSGEAPNOAHILKETEFKVKVLSGAFGTYKGLMPBGEKVIPIALILEB 173
QY 763 ENTSPKANKETLDEAVVAGVSGPYVSRLLGICLTSTVOLVQLMPYGCCLLDHVENRGR 822
DB 174 EATSPKANKETLDEAVVAGVSDPNHVCRLIGICLTSTVOLITQMPYGCCLLDYREHKDN 233
QY 823 LGSODLNMCMQIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITPFGILRLIDETET 882
DB 234 IGSQYLLNMCMQIAKGMNLYEERLVRDLAARVLYKTPQHVKITDFGLAKLGADKE 293
QY 883 YHADGKVPIMKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGI-PAREIPDL 942
DB 294 YHADGKVPIMKMALESILRRITTHOSDWSYGVTVWELMTFGSKPYDGI-PASEISSV 353
QY 943 KGERLPQRPITIDVYIMVCKMMDISECRPRRELVSERPMARDPQRFVIO-NEDLG 1001
DB 354 KGERLPQRPITIDVYIMVCKMMDIDSRPKFRELIAPSKMARDPRLVIOGDERMH 413
QY 1002 PASPLDSTFYRSLLEDDMDGLVDAERYLVPOGFCFDPAPAGAGMTHHRSSSTRSG 1061
DB 414 LPSPFDSKFRITLMEEDMEDIVDADEYLVPHOGFF-----NSPST--- 454
QY 1062 GGDUTLGLPSEEARSPPL-----APSEGASDVFPDGLQMGAKGLQSLPTHPD 1116
DB 455 -----SRPLLSLSLATSNNASATNCIDRNG-----H----- 481
QY 1117 RYSEDPLVPPLSERIDGYVAPLTCSPQPEYVNOQPVRRPQSPREGPLPAPAPAGAT-LER 1175
DB 482 -----PVREDGFL-----PAPRYVNO--LMPKKSTAMVONOYVYISLTAISK 523
QY 1176 AKTLSPKNGVVKDVFAPGAVENPEYL 1203
DB 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987
 A>Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
 A:Reference number: S00727; MUID:88217326; PMID:2897102
 A:Accession: S00727
 A:Molecule type: DNA
 A:Residues: 1-545 <SCO>
 A:Cross-references: EMBL:X06943
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; phosphotransferase
 F:135-400/Domain: protein kinase homology <KIN>
 F:133-151/Region: protein kinase ATP-binding motif

Query Match 24.0%; Score 1640; DB 2; Length 545;
 Best Local Similarity 54.9%; Pred. No. 2.9e-60;
 Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 584 GREDDCVACHYDPPFCVACRSPGKPDLSYPIKFPDEBAGQPCPINCTHSCVDL 643
 DB 1 GP--DHCMKCAHFIDGPHCVACRAGVIGENDTL-VMKYADANAVCQLCHPCTRGCKGP 57
 QY 644 DDGCPABQASPLTISVSAVY-GILLVYVGVVFGILIKRQOKIRKTYMRLLQETEL 702
 DB 58 GLEBCP--NGSKTPSIAAGVVGGLCLVVGGLGILYLR--HIVRRTRLRLQEREL 113
 QY 703 VEPPLPSGAMPNQOMRILKETELRKVVLGSGAFYVKGIMIPDGENVKIPVAIKYLR 762
 DB 114 VEPPLPSGAMPNQOMRILKETELRKVVLGSGAFYVKGIMIPDGENVKIPVAIKYLR 173
 QY 763 ENTPSKANKELDEAYYVAGVSPYVSRLLGICLTSTVQVLTQIMPYGCLLDHVRNNGR 822
 DB 174 EATSPKANKELDEAYYVAGVSPYVSRLLGICLTSTVQVLTQIMPYGCLLDHVRNNGR 233
 QY 823 LGSODLLNMCQIAGKMSYLEDVRLVHRDLAARVNVKSPKRVKITTDEGLARLLDIDTE 882
 DB 234 IGSODLLNMCQIAGKMSYLEDVRLVHRDLAARVNVKSPKRVKITTDEGLARLLDIDTE 293
 QY 883 YHADGKVPKIMWALLESILRRFTHOSDVSYGTVWELMTFGAKPYDGIPIARELPDLE 942
 DB 294 YHADGKVPKIMWALLESILRRFTHOSDVSYGTVWELMTFGAKPYDGIPIARELPDLE 353
 QY 943 KGERLPDPPICTIDVYIMVYKCMWIDSECRPREFELVSEFSRMAPDPQRFVVO--NEDLG 1001
 DB 354 KGERLPDPPICTIDVYIMVYKCMWIDSECRPREFELVSEFSRMAPDPQRFVVO--NEDLG 413
 QY 1002 PASPLDSTFYRSLLEDMDGLVDAREYLVPOQGFPCDPAPRAGAGMWHNRSSSTRSG 1061
 DB 414 LPSPTDSKFRITLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
 QY 1062 GGDLLGLPSESEAPRSP-----APSEGASDVFPDGLGMAKGLQSLPTDPSPLQ 1116
 DB 455 -----SKTPLLSSLSATSNNSATNCIDRNG-----H----- 481
 QY 1117 RYSEDPVLPSETDGVAPLTCSPQPEYVNOQPVPRQPSPREGRLPAARPAAGAT--LER 1175
 DB 482 -----PYREDDGL-----PAPEYVNO--LMPKKRSTAMVQVIVYISLTATISK 523
 QY 1176 AKTLSPKNGVNVQVDFAFGAVENPEYL 1203
 DB 524 LPMDSRYON-----SHSTAVDNPPEYL 544

RESULT 15

GQFFE

epidermal growth factor receptor - fruit fly (Drosophila melanogaster)

N:Contains: protein-tyrosine kinase (BC 2.7.1.112) erbB

C:Species: Drosophila melanogaster

C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999

C:Accession: A00640; A38021

R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B. Z.

Cell 40, 599-607, 1985

A>Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding

A:Reference number: A00640; MUID:85124611; PMID:2982499
 A:Accession: A00640
 A:Molecule type: DNA
 A:Residues: 1-1330 <LIV>
 A:Cross-references: EMBL:K03054
 R:Madsworth, S. C.; Vincent III, W. S.; Bildeau-Wentworth, D.
 Nature 314, 178-180, 1985
 A>Title: A Drosophila genomic sequence with homology to human epidermal growth factor
 A:Reference number: A38021; MUID:85137938; PMID:2983232
 A:Accession: A38021

A:Molecule type: DNA
 A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
 A:Cross-references: EMBL:X02293; NID:67922; PIDN:CAA26157.1; PID:929565
 C:Comment: This sequence is tentative because the introns have not been identified.

A:Gene: FlyBase: Egfr
 A:Cross-references: FlyBase: FBgn0003731

A:Map position: 2 57F

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph

F:1-732/Domain: extracellular #status predicted <EXT>

F:733-764/Domain: transmembrane #status predicted <TM>

F:765-1330/Domain: intracellular #status predicted <INT>

F:808-1072/Domain: protein kinase homology <KIN>

F:816-824/Region: protein kinase ATP-binding motif

F:1122-300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status

F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F:843/Active site: lys #status predicted

F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predi

Query Match 23.8%; Score 1626.5; DB 1; Length 1330;
 Best Local Similarity 30.2%; Pred. No. 2.4e-59;
 Matches 418; Conservative 179; Mismatches 411; Indels 377; Gaps 45;

QY 80 VQGVYLAHNGVROVPLQRLRIVRGTOF-----EDNYALVLDNGDPLNTPVTGASP 134
 DB 38 ITNYIVIGLDIRPCTLSYRLQIRGRITFLSVSEBEKALFY-----TY 81
 QY 135 GGLRELQRLSTELLKGVYLQRPOLCYOTILMKDIFHKQNALTLIDNRSRACHP 194
 DB 82 SKMTLETPDLRDVLNQGVRHNNYMLCHMTQIMSEIVNSGTAAYVYDFAPAREPRCP 141
 QY 195 CSPMKGRSCWGESSEDDQSLRTVYACAGCA--RCKGRLPTDCCHEOGAAGTGPKHSDC 252
 DB 142 CHESTG--CMGEPRKQOKSKLTCSPOCGRGKYPKPCCHLFCAGGCTGTQKDC 200
 QY 253 LACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCYVACPYNYLSTVGSCT 312
 DB 201 IACNFFDEAVSKCECPMRKYNTPTTYLLETPNSGKYAYGATVKECP--CHLLRNGACV 259
 QY 313 LVCPYLNQVYAEVGTOCEKSKPCARVCYGLGMOYIKANSKFIGITEL-----EPAG 366
 DB 260 RSCPQDRKDKGE-----CVPCNGPRCPKTC-----PGYVTLAHGINDSPRN 300
 QY 367 CKKIFGSLAFIPESFDG--DPASNTA-----PLQRPQVFEETLEETGYLISAWMDS 418
 DB 301 CTVIDGNIRILDQTFSGFDVYANYTNGPRYTPLDPEKREVFSTVKEITGYLNEGTNPQ 360
 QY 419 LPDLVFPQNLQVIRGRIIHNGAY--SLTLQIGISMLGRLSRLREAGSLALIHNTTLGFV 477
 DB 361 FRNLSTFRNLLETINGRLQMSMFALAIIVSSLSLEMRNKKQISSGSVIQRNRDLCYV 420
 QY 478 HTVPW-----DQLFRNNFTVS----- 494
 DB 421 SNIRMPAIQKPEBKVVNENLRADLGKFTLLISVOHNIIMHIFALCREKMHLLGSV 480
 QY 495 -----FMRLV-----PKYSA 504
 DB 481 QRGLLGSMTGVSVPYLOELOFQNHRLRLWLYLOVINSIODKSNHQLTDACYSVPVT 540
 QY 505 S-----HLEE-----C-----VG 512
 DB 541 SLTERARVALQSGALMELQITARASMRHSKTLRABGRQVRWVGLVCASABAGIA 600

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842
1 MELALCRWGLLLALLPPQA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6626	96.8	1255	1	ERB2_HUMAN
2	5852	85.5	1257	1	ERB2_RAT
3	5846.5	85.5	1254	1	ERB2_MESAU
4	3116	45.5	1210	1	EGFR_HUMAN
5	3087	45.1	1210	1	EGFR_MOUSE
6	2942.5	43.0	1308	1	ERB4_HUMAN
7	2924	42.7	1308	1	ERB4_RAT
8	2677.5	39.1	1167	1	XMRK_XIPMA
9	2390.5	34.9	1342	1	ERB3_HUMAN
10	2323.5	34.0	1339	1	ERB3_RAT
11	1920	28.1	1426	1	EGFR_DROME
12	1749.5	25.6	634	1	ERBB_ALV
13	1703	24.9	604	1	ERBB_AVIER
14	1630	23.8	540	1	ERBB_AVIER
15	1571	23.0	703	1	EGFR_CHICK
16	1276	18.6	1323	1	L723_CAEL
17	1142.5	16.7	245	1	ERB2_MOUSE
18	717	10.5	1363	1	ILPR_BRALA
19	702	10.3	1383	1	INSR_RAT
20	701.5	10.3	1372	1	INSR_MOUSE
21	701	10.2	1300	1	INSR_MOUSE
22	701	10.2	1382	1	INSR_HUMAN
23	693	10.1	1297	1	IRK_HUMAN
24	691	10.1	1477	1	HTK7_HYDAT
25	684.5	10.0	1300	1	IRR_CAVPO
26	680	9.9	1607	1	MIPR_LYMT
27	645	9.4	1367	1	IGR_HUMAN
28	630	9.2	1390	1	INSR_AEDAE
29	626	9.1	1373	1	IGR_MOUSE
30	623	9.1	2146	1	INSR_DROME
31	622.5	9.1	1370	1	IGR_RAT
32	601	8.8	987	1	EPB4_HUMAN
33	597.5	8.7	984	1	EPB1_CHICK

34	597.5	8.7	984	1	EPB1_RAT	P09759	rattus norv
35	591.5	8.6	984	1	EPB1_HUMAN	P54762	homo sapien
36	587.5	8.6	977	1	EPB2_MOUSE	Q03145	mus musculu
37	586	8.6	902	1	EPBB_XENLA	Q91736	xenopus lae
38	585.5	8.6	976	1	EPB2_HUMAN	P29317	homo sapien
39	583	8.5	987	1	EPB4_MOUSE	P54761	mus musculu
40	578	8.4	1114	1	RET_HUMAN	P07949	homo sapien
41	573.5	8.4	1053	1	PAK1_CHICK	Q00944	gallus gall
42	572.5	8.4	985	1	EPBA_XENLA	Q91571	xenopus lae
43	569	8.3	1068	1	FAT1_XENLA	Q91736	xenopus lae
44	565.5	8.3	757	1	HT16_HYDAT	P53356	hydra atten
45	565.5	8.3	986	1	EPBA_XENLA	Q91845	xenopus lae

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 41, Last annotation update)
DT 15-JUN-2002 (Rel. 05, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=86119663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Salto T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234 (1986).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139 (1985).
RN (3)
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).
RN (4)
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429 (1993).
RN (5)
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
RN ALPHA AND AMPHIREGULIN.
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -!- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -!- ALLELE TYPE: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M11767; AAA35808.1; -
 CC EMBL: M11761; AAA35808.1; JOINED.
 CC EMBL: M11762; AAA35808.1; JOINED.
 CC EMBL: M11763; AAA35808.1; JOINED.
 CC EMBL: M11764; AAA35808.1; JOINED.
 CC EMBL: M11765; AAA35808.1; JOINED.
 CC EMBL: M11766; AAA35808.1; JOINED.
 CC EMBL: M11730; AAA35808.1; -
 CC EMBL: M12036; AAA35978.1; -
 CC EMBL: X01363; CAA27060.1; -
 CC PIR: A25491; A25491.
 CC PIR: A24571; A24571.
 CC HSSP: P11362; 1FGK.
 CC DR Genew; HGNC:3430; ERBB2.
 CC DR MIM: 164870; -
 CC DR InterPro; IPR000494; EGFR_L_domain.
 CC DR InterPro; IPR000719; Euk_pkinase.
 CC DR InterPro; IPR002174; Furin-like.
 CC DR InterPro; IPR001245; Tyr_pkinase.
 CC DR InterPro; IPR004019; YLP_motif.
 CC DR Pfam; PF00069; pkinase; 1.
 CC DR Pfam; PF00757; Furin-like; 1.
 CC DR Pfam; PF01030; Recep_L_domain; 2.
 CC DR Pfam; PF02757; YLP; 2.
 CC DR Prodom; PD000001; Euk_pkinase; 1.
 CC DR SMART; SMO0261; Fy; 3.
 CC DR SMART; SMO0219; Tyrc; 1.
 CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 CC KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC KW Polymorphism.
 CC KW SIGNAL.
 CC FT CHAIN 1 21 POTENTIAL.
 CC FT DOMAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 CC FT TRANSMEM 653 652 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 676 1255 POTENTIAL.
 CC FT DOMAIN 720 987 CYTOPLASMIC (POTENTIAL).
 CC FT NP_BIND 726 734 PROTEIN KINASE.
 CC FT BINDING 753 753 ATP (BY SIMILARITY).
 CC FT ACT_SITE 845 845 ATP (BY SIMILARITY).
 CC FT DISULFID 195 204 BY SIMILARITY.
 CC FT DISULFID 199 212 BY SIMILARITY.
 CC FT DISULFID 220 227 BY SIMILARITY.
 CC FT DISULFID 224 235 BY SIMILARITY.
 CC FT DISULFID 236 244 BY SIMILARITY.
 CC FT DISULFID 240 252 BY SIMILARITY.
 CC FT DISULFID 255 264 BY SIMILARITY.
 CC FT DISULFID 268 295 BY SIMILARITY.
 CC FT DISULFID 299 311 BY SIMILARITY.
 CC FT DISULFID 315 331 BY SIMILARITY.
 CC FT DISULFID 334 338 BY SIMILARITY.
 CC FT DISULFID 511 520 BY SIMILARITY.
 CC FT DISULFID 515 528 BY SIMILARITY.
 CC FT DISULFID 531 540 BY SIMILARITY.
 CC FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 584 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 /FTID=VAR_004077.
 FT VARIANT 655 655 I -> V.
 FT VARIANT 655 655 /FTID=VAR_004078.
 FT CONFLICT 1170 1170 P -> A (IN REF. 2).
 FT SEQUENCE 1255 AA; 137909 MW; 399DFDA04DCF962 CRC64;
 SQ
 Query Match 96.8%; Score 6626; DB 1; Length 1255;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
 QY 1 MELALCRWGLLLALLPGAASTOVCTGDMKRLRASPETHLDMLRHLVYGGQVQGNL 60
 DB 1 MELALCRWGLLLALLPGAASTOVCTGDMKRLRASPETHLDMLRHLVYGGQVQGNL 60
 QY 61 ELTYLPTASISFLDIOGVGVLIANOVQVPLQRIIRYRGTLFEDNVALVLDNG 120
 DB 61 ELTYLPTASISFLDIOGVGVLIANOVQVPLQRIIRYRGTLFEDNVALVLDNG 120
 QY 121 DELNNTTPTVGASPGGLRELQRLSTEILKGVLIQRLNQLQYDTILKKDIFKKNOLA 180
 DB 121 DELNNTTPTVGASPGGLRELQRLSTEILKGVLIQRLNQLQYDTILKKDIFKKNOLA 180
 QY 122 DELNNTTPTVGASPGGLRELQRLSTEILKGVLIQRLNQLQYDTILKKDIFKKNOLA 180
 DB 122 DELNNTTPTVGASPGGLRELQRLSTEILKGVLIQRLNQLQYDTILKKDIFKKNOLA 180
 QY 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRVYACAGCARCKPLPTDCHEOC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRVYACAGCARCKPLPTDCHEOC 240
 QY 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRVYACAGCARCKPLPTDCHEOC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRVYACAGCARCKPLPTDCHEOC 240
 QY 241 AAGCTGPRKSDCLACHFNHSGICELHPCALVTYNTDTSMPNPGRTTFGASCYACP 300
 DB 241 AAGCTGPRKSDCLACHFNHSGICELHPCALVTYNTDTSMPNPGRTTFGASCYACP 300
 QY 241 AAGCTGPRKSDCLACHFNHSGICELHPCALVTYNTDTSMPNPGRTTFGASCYACP 300
 DB 241 AAGCTGPRKSDCLACHFNHSGICELHPCALVTYNTDTSMPNPGRTTFGASCYACP 300
 QY 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSPCARVCYGLCMOYIKANSKFI 360
 DB 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSPCARVCYGLCMOYIKANSKFI 360
 QY 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSPCARVCYGLCMOYIKANSKFI 360
 DB 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSPCARVCYGLCMOYIKANSKFI 360
 QY 361 ELEPAGCKKIFGSLAPLPESPDGDPASNTAPLQPEQLVFTLEBITGYLYISAWD 420
 DB 361 ELEPAGCKKIFGSLAPLPESPDGDPASNTAPLQPEQLVFTLEBITGYLYISAWD 420
 QY 361 ELEPAGCKKIFGSLAPLPESPDGDPASNTAPLQPEQLVFTLEBITGYLYISAWD 420
 DB 361 ELEPAGCKKIFGSLAPLPESPDGDPASNTAPLQPEQLVFTLEBITGYLYISAWD 420
 QY 421 DLSVONLQVIRGRLHNGAVSLTQIGLISWGLRSLRELSSGLALHHNTHLCVHY 480
 DB 421 DLSVONLQVIRGRLHNGAVSLTQIGLISWGLRSLRELSSGLALHHNTHLCVHY 480
 QY 421 DLSVONLQVIRGRLHNGAVSLTQIGLISWGLRSLRELSSGLALHHNTHLCVHY 480
 DB 421 DLSVONLQVIRGRLHNGAVSLTQIGLISWGLRSLRELSSGLALHHNTHLCVHY 480
 QY 481 PMDOLFRFNFTVSFMLRVKVSASHLECEVGEGLACHOLCARGHGMCGRPTCVNCSOF 540
 DB 481 PMDOLFRFNFTVSFMLRVKVSASHLECEVGEGLACHOLCARGHGMCGRPTCVNCSOF 540
 QY 481 PMDOLFRFNFTVSFMLRVKVSASHLECEVGEGLACHOLCARGHGMCGRPTCVNCSOF 540
 DB 481 PMDOLFRFNFTVSFMLRVKVSASHLECEVGEGLACHOLCARGHGMCGRPTCVNCSOF 540
 QY 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGPADOCVACAHYKDP 600
 DB 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGPADOCVACAHYKDP 600
 QY 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGPADOCVACAHYKDP 600
 DB 541 LRGECEVECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGPADOCVACAHYKDP 600
 QY 601 PCVACPSGVKPDLSYMPYWKPRDEEGACOPRINCHSCVLDKGCAGAEQASPLTSI 660
 DB 601 PCVACPSGVKPDLSYMPYWKPRDEEGACOPRINCHSCVLDKGCAGAEQASPLTSI 660
 QY 601 PCVACPSGVKPDLSYMPYWKPRDEEGACOPRINCHSCVLDKGCAGAEQASPLTSI 660
 DB 601 PCVACPSGVKPDLSYMPYWKPRDEEGACOPRINCHSCVLDKGCAGAEQASPLTSI 660
 QY 661 VSAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI 720
 DB 661 VSAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI 720
 QY 661 VSAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI 720
 DB 661 VSAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI 720
 QY 655 ISAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI 714
 DB 655 ISAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAPNQAOMRI 714

```

QY 721 LKTELARKVVLGSGAGCTYKGIWIDGENVKIPVAIKVIRENTSPANKNEIDEAYVM 780
Db 715 LKTELARKVVLGSGAGCTYKGIWIDGENVKIPVAIKVIRENTSPANKNEIDEAYVM 774
QY 781 AGVSPVYSRLGICLTSTVOLVQLMPYGLLDHVRNRRRLSODLNNCMQIAKMS 840
Db 775 AGVSPVYSRLGICLTSTVOLVQLMPYGLLDHVRNRRRLSODLNNCMQIAKMS 834
QY 841 YLEDLRVLRDLAARNLVKSPNHVKTITDFGLARLLDIDETEHADGKVPDKMMALESI 900
Db 835 YLEDLRVLRDLAARNLVKSPNHVKTITDFGLARLLDIDETEHADGKVPDKMMALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLLEKGERLPQPEICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLLEKGERLPQPEICTIDVYMI 954
QY 961 MVKCMKIDSECRPFRELVSFSSRMARDPQRFVVIQNEIDLQPSAPLQDSTFYRSLLEDDDM 1020
Db 955 MVKCMKIDSECRPFRELVSFSSRMARDPQRFVVIQNEIDLQPSAPLQDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEEYLVPOQGFCCPDPAAGGVVHRRHSSSTRSGCDLTLGLPSEEEAPRSP 1080
Db 1015 GDLVDAEEYLVPOQGFCCPDPAAGGVVHRRHSSSTRSGCDLTLGLPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAKGICLSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAKGICLSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1134
QY 1141 POPPYVQOPVPRPQSPRPSGRLPAARPAATLERAKTSLGKGVVADVAFGGAENP 1200
Db 1135 POPPYVQOPVPRPQSPRPSGRLPAARPAATLERAKTSLGKGVVADVAFGGAENP 1194
QY 1201 EYLTPQCGAAPQHPPPAFSPADNLVYMDQDPERGAPSTPFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPQCGAAPQHPPPAFSPADNLVYMDQDPERGAPSTPFKGTPTAENPEYGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

```

```

RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loties F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the new protein.";
RL EMBL J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03362; CAA27059.1; ALU_INIT.
DR PIR; A24562; TYR2NU.
DR HSSP; P1362; 1EGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_Like; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 1 21
FT DOMAIN 22 1257
FT DOMAIN 22 654
FT TRANSMEM 655 677
FT DOMAIN 678 1257
FT DOMAIN 159 369
FT DOMAIN 473 646
FT DOMAIN 722 989
FT NP_BIND 728 736
FT BINDING 755 755
FT ACT_SITE 847 847
FT DISULFID 196 205
FT DISULFID 200 213
FT DISULFID 221 228
FT DISULFID 225 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 312
FT DISULFID 316 332
FT DISULFID 335 339
FT DISULFID 513 522

```

```

FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT MOD_RSS 632 644 BY SIMILARITY.
FT MOD_RSS 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.5%; Score 5852; DB 1; Length 1257;
Best Local Similarity 85.8%; Pred. No. 1.6e-305;
Matches 1085; Conservative 52; Mismatches 117; Indels 10; Gaps 3;

QY 1 MELALGRWGLLLLPFGAASVCTGTDKRLPASPTHLMDRLHYOGCQVQGNL 60
D 1 MELALGRWGLLLLPFGAASVCTGTDKRLPASPTHLMDRLHYOGCQVQGNL 60
QY 61 ELTYLPNNAISLFIODIOEVQVYLIANOVROVPLQRLRYRGTOLEFEDNYALVDNG 120
D 61 ELTYLPNNAISLFIODIOEVQVYLIANOVROVPLQRLRYRGTOLEFEDNYALVDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQSLTEILKGVLIQNPQLCYODITLMDKIDPHKNOL 179
D 121 DPLNNTTPVT-GASPGGLRELQSLTEILKGVLIQNPQLCYODITLMDKIDPHKNOL 179
QY 121 DPQNVAASTFGRTPEGLRELQSLTEILKGVLIQNPQLCYODITLMDKIDPHKNOL 180
D 121 DPQNVAASTFGRTPEGLRELQSLTEILKGVLIQNPQLCYODITLMDKIDPHKNOL 180
QY 180 ALTLIDNRSRACHPCSPKMGSRKMGWSESSDCSLTRTVAGGACGACRCKPLPTDCHEQ 239
D 180 ALTLIDNRSRACHPCSPKMGSRKMGWSESSDCSLTRTVAGGACGACRCKPLPTDCHEQ 239
QY 181 APVIDNRSRACHPCSPKMGSRKMGWSESSDCSLTRTVAGGACGACRCKPLPTDCHEQ 240
D 181 APVIDNRSRACHPCSPKMGSRKMGWSESSDCSLTRTVAGGACGACRCKPLPTDCHEQ 240
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTPESMPNBEGRYTFGASCVTAC 299
D 240 CAAGCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTPESMPNBEGRYTFGASCVTAC 299
QY 241 CAAGCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTPESMPNBEGRYTFGASCVTAC 300
D 241 CAAGCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTPESMPNBEGRYTFGASCVTAC 300
QY 300 PYNVLSTDVGCTVCPPLHNOEVAEDTORCEKSKCARVCGYLGMOYIKANSKFTGI 359
D 300 PYNVLSTDVGCTVCPPLHNOEVAEDTORCEKSKCARVCGYLGMOYIKANSKFTGI 359
QY 301 PYNVLSTDVGCTVCPPLHNOEVAEDTORCEKSKCARVCGYLGMOYIKANSKFTGI 360
D 301 PYNVLSTDVGCTVCPPLHNOEVAEDTORCEKSKCARVCGYLGMOYIKANSKFTGI 360
QY 360 TELEFAGCKKIFGSLAPLPSFDDPASNTAPLQPEQLQVELEIEITGYLISAMPDNL 419
D 360 TELEFAGCKKIFGSLAPLPSFDDPASNTAPLQPEQLQVELEIEITGYLISAMPDNL 419
QY 361 NVGFQDCKKIFGSLAPLPSFDDPASNTAPLQPEQLQVELEIEITGYLISAMPDNL 420
D 361 NVGFQDCKKIFGSLAPLPSFDDPASNTAPLQPEQLQVELEIEITGYLISAMPDNL 420
QY 420 PDLVSFONLOVIRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 479
D 420 PDLVSFONLOVIRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 479
QY 421 RDLVSFONLOVIRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 480
D 421 RDLVSFONLOVIRGRILHNGAYSLTLOGLISWLGRLRELGSGLAIHHNTHLCPVHT 480
QY 480 VPMDQLFRFNFTVSPMLRVKVSASHLEE--CYGEGLAGHQLCARGHGCMGSPPTQCN 537
D 480 VPMDQLFRFNFTVSPMLRVKVSASHLEE--CYGEGLAGHQLCARGHGCMGSPPTQCN 537
QY 481 VPMDQLFRFNFTVSPMLRVKVSASHLEE--CYGEGLAGHQLCARGHGCMGSPPTQCN 538
D 481 VPMDQLFRFNFTVSPMLRVKVSASHLEE--CYGEGLAGHQLCARGHGCMGSPPTQCN 538
QY 538 SGLRGECEVEECVLOGLPREYVNAHRLCHPECOPOGNSVTCFGEADQCAVACAYK 597
D 538 SGLRGECEVEECVLOGLPREYVNAHRLCHPECOPOGNSVTCFGEADQCAVACAYK 597
QY 534 SHFLRGECEVEECVLOGLPREYVNAHRLCHPECOPOGNSVTCFGEADQCAVACAYK 593
D 534 SHFLRGECEVEECVLOGLPREYVNAHRLCHPECOPOGNSVTCFGEADQCAVACAYK 593
QY 598 DPPPCVACRSGVVPDLSYPIKFPPEEGACQPCPINCHTSCVDLDKGPAPQAPQAPL 657
D 598 DPPPCVACRSGVVPDLSYPIKFPPEEGACQPCPINCHTSCVDLDKGPAPQAPQAPL 657
QY 594 DSSSCVACRSGVVPDLSYPIKFPPEEGACQPCPINCHTSCVDLDKGPAPQAPQAPL 653
D 594 DSSSCVACRSGVVPDLSYPIKFPPEEGACQPCPINCHTSCVDLDKGPAPQAPQAPL 653
QY 658 TSIYSAVAGLILVVVLGVFGILIKRROOKIRKTYMRLIODETELYVEPIYSGAMPNQAQ 717
D 658 TSIYSAVAGLILVVVLGVFGILIKRROOKIRKTYMRLIODETELYVEPIYSGAMPNQAQ 717
QY 654 TFIATYVGVILFLLVVVGVILIKRRROOKIRKTYMRLIODETELYVEPIYSGAMPNQAQ 713
D 654 TFIATYVGVILFLLVVVGVILIKRRROOKIRKTYMRLIODETELYVEPIYSGAMPNQAQ 713
QY 718 MRIKETELRKVKVLGSGAGFTYVKGWIPDGENVKIPVALIKVLRNTPRANKKILDEA 777
D 718 MRIKETELRKVKVLGSGAGFTYVKGWIPDGENVKIPVALIKVLRNTPRANKKILDEA 777

```

```

D 714 MRIKETELRKVKVLGSGAGFTYVKGWIPDGENVKIPVALIKVLRNTPRANKKILDEA 773
QY 778 YVMAVGSPYVRLIGTICLSTVOLVTOAMPYGLLDHVENRGRGSGODLNMCMQIAK 837
D 778 YVMAVGSPYVRLIGTICLSTVOLVTOAMPYGLLDHVENRGRGSGODLNMCMQIAK 837
QY 774 YVMAVGSPYVRLIGTICLSTVOLVTOAMPYGLLDHVENRGRGSGODLNMCMQIAK 833
D 774 YVMAVGSPYVRLIGTICLSTVOLVTOAMPYGLLDHVENRGRGSGODLNMCMQIAK 833
QY 838 GMSVLEPDLVRLHRLAARNVYVKSNNHYKINDPGIARLDDIDETEVHADGGVPIKMMAL 897
D 838 GMSVLEPDLVRLHRLAARNVYVKSNNHYKINDPGIARLDDIDETEVHADGGVPIKMMAL 897
QY 834 GMSVLEPDLVRLHRLAARNVYVKSNNHYKINDPGIARLDDIDETEVHADGGVPIKMMAL 893
D 834 GMSVLEPDLVRLHRLAARNVYVKSNNHYKINDPGIARLDDIDETEVHADGGVPIKMMAL 893
QY 898 ESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRLLPQPICTIDV 957
D 898 ESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRLLPQPICTIDV 957
QY 894 ESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRLLPQPICTIDV 953
D 894 ESILRRRFTTHSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGRLLPQPICTIDV 953
QY 958 YMIWKCMMIDSECRPREFRELVESESRMARPPORFVVIQNEIDLGPASPLDSTFYSLLED 1017
D 958 YMIWKCMMIDSECRPREFRELVESESRMARPPORFVVIQNEIDLGPASPLDSTFYSLLED 1017
QY 954 YMIWKCMMIDSECRPREFRELVESESRMARPPORFVVIQNEIDLGPASPLDSTFYSLLED 1013
D 954 YMIWKCMMIDSECRPREFRELVESESRMARPPORFVVIQNEIDLGPASPLDSTFYSLLED 1013
QY 1018 DDMGDLDVAEIEYLPQOGFCPPDPAPAGGAVHHRHSSSTRSGGDLTLGLEPSEEAR 1077
D 1018 DDMGDLDVAEIEYLPQOGFCPPDPAPAGGAVHHRHSSSTRSGGDLTLGLEPSEEAR 1077
QY 1014 DDMGDLDVAEIEYLPQOGFCPPDPAPAGGAVHHRHSSSTRSGGDLTLGLEPSEEAR 1073
D 1014 DDMGDLDVAEIEYLPQOGFCPPDPAPAGGAVHHRHSSSTRSGGDLTLGLEPSEEAR 1073
QY 1078 RSLPAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVAPL 1137
D 1078 RSLPAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVAPL 1137
QY 1074 RSLPAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVAPL 1133
D 1074 RSLPAPSEAGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVAPL 1133
QY 1138 TCSPOPEYVNPQDVPPPSRPGPLPAAPAGATLEBAKTLSPGKNGVXKDVFAFGAV 1197
D 1138 TCSPOPEYVNPQDVPPPSRPGPLPAAPAGATLEBAKTLSPGKNGVXKDVFAFGAV 1197
QY 1134 ACSPOPEYVNPQDVPPPSRPGPLPAAPAGATLEBAKTLSPGKNGVXKDVFAFGAV 1193
D 1134 ACSPOPEYVNPQDVPPPSRPGPLPAAPAGATLEBAKTLSPGKNGVXKDVFAFGAV 1193
QY 1198 ENPEYLPVREGTASPPHSPAPFADNLVYWDODPERGAPSPFFKTPAENPEYGL 1257
D 1198 ENPEYLPVREGTASPPHSPAPFADNLVYWDODPERGAPSPFFKTPAENPEYGL 1257
QY 1194 ENPEYLPVREGTASPPHSPAPFADNLVYWDODPERGAPSPFFKTPAENPEYGL 1253
D 1194 ENPEYLPVREGTASPPHSPAPFADNLVYWDODPERGAPSPFFKTPAENPEYGL 1253
QY 1258 DVPV 1261
D 1258 DVPV 1257
QY 1254 DVPV 1257
D 1254 DVPV 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
AC 06053;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
DN ERB2 OR NEU.
OS Mesocricetus auratus (golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

```

CC RESIDUES.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D16295; BAA03801.1; -
 DR HSSP: P11362; IFGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM0026; FU; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 KM Transmembrane, Glycoprotein; Multigene family; Receptor; Signal;
 KM transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM proto-oncogene; Disease mutation.
 FT CHAIN 1 21
 FT SIGNAL 1 21
 FT CHAIN 22 1254
 FT DOMAIN 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1254
 FT DOMAIN 158 368
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP BIND 726 734
 FT BINDING 753 753
 FT ACT SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560
 FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 630 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT CARBOHYD 658 658
 FT VARIANT 659 659
 FT VARIANT 1254 AA; 138252 KM; 974C3791C21F2B1 CRC64;
 SQ SEQUENCE

Query Match 85.5%; Score 5846.5; DB 1; Length 1254;
 Best Local Similarity 85.3%; Pred. No. 3.2e-305;
 Matches 1080; Conservative 60; Mismatches 109; Indels 17; Gaps 3;
 QY 1 MELALCRWGLLALLPGAASTGYCTGDMKRLPASETHLMDRLHYGCGVQVGNL 60
 DB 1 MELAMCGGLLALLSPASGTYCTGDMKRLPASETHLMDRLHYGCGVQVGNL 60
 QY 61 ELTYVPTNASLSPLODIOGVGVLLAHNOVQVPLQRLIRYRGTOLEFDNVALATLNDG 120
 DB 61 ELTYVPANATLSPLDDIOGVGVMLAHQVHRVPLQRLIRYRGTOLEFDKVALAVALDNR 120
 QY 121 DPLNNTPTVGTASPGGLRELQRLSTEILKGVLLIORNPQLCYODITLMMKDIPHKNOOLA 180
 DB 121 DPLDNTTATGRTPEGLRELQRLSTEILKGVLLIRGNQLCYODITLMMKDVRKNOOLA 180
 QY 181 LTLIDTNRSRACHPCSPMKGSRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
 DB 181 PVDIDTNRSRACPPCAPACKDNHCWASPEDCOTLGTIAPRAVAPARARLPDCCHEQC 240
 QY 241 AAGCTGPRKSDCLACLFHNSGICELHCPALVTYNTDTRESNPNREGRTTFGASCTYACP 300
 DB 241 AAGCTGPRKSDCLACLFHNSGICELHCPALVTYNTDTRESNPNREGRTTFGASCTYACP 300
 QY 301 VYVLTSDVGSCTLVCPPLHNOEYTAEDGTORCEKSPCARVCYGLMOYIKANSKFIGIT 360
 DB 301 VYVLTSDVGSCTLVCPPLHNOEYTAEDGTORCEKSPCARVCYGLMOYIKANSKFIGIT 360
 QY 361 ELEFAGCKKIFGSLAFLEPSPDGPASNTAPLOPELQVFEETLEETIGVLYISAMPDLSL 420
 DB 361 IDEFAGCKKIFGSLAFLEPSPDGPASNTAPLOPELQVFEETLEETIGVLYISAMPDLSL 420
 QY 421 DLSVQNTQVYGRILNCAVSLTQGLISWGLRSLRELSCGLAIHHNTHLCFVHTV 480
 DB 421 DLSVQNTQVYGRILNCAVSLTQGLISWGLRSLRELSCGLAIHHNTHLCFVHTV 480
 QY 481 PMDOLFRRNPFVFLRPVKSASHL-----EECGEGLACHOCARHCWGPGTQCV 535
 DB 481 PMDOLF-----RNPHQALHSGNPSEEGGLKDPACYPICAHCHWGPGTQCV 535
 QY 536 NCSQPLRQGECEVEGRVLOGLPREYVNAHCLPCHPECPONGSVTCFPEADQCVACAH 595
 DB 536 NCSHRLRQGECEVEGRVLOGLPREYVNGHCLPCHPECPONSTETCTSEADQCTACH 589
 QY 596 YKDPPECVACRCPGKPDLSYMPWKFPDEBACQPCPICTHSCVYDLDDKCPABQRAS 655
 DB 596 YKDSPEFCVACRCPGKPDLSYMPWKFPDEBACQPCPICTHSCVYDLDERGCPABQRAS 649
 QY 656 PLTSVSAVNGILLVVLGVNFGILLIKRQOKIRKTYARLLQETELVPLTPSGAMPNO 715
 DB 656 PATSIITATVGLLFLVIGVAVGILLIKRRQOKIRKTYARLLQETELVPLTPSGAMPNO 709
 QY 716 AQMRILKETELRKVAVLGSAGAGTYKGIWIPDGENVKIPVAIKVLENTSPKANEILD 775
 DB 716 AQMRILKETELRKVAVLGSAGAGTYKGIWIPDGENVKIPVAIKVLENTSPKANEILD 769
 QY 776 EAYVAVGSPYVSSLGICLTSTVQLVTLMPYCLLDHVENRGRGLSODLNNCMQI 835
 DB 776 EAYVAVGSPYVSSLGICLTSTVQLVTLMPYCLLDHVENRGRGLSODLNNCMQI 829
 QY 836 AKGMSYLEDVRLVHNDLAARNVLVSPNHNKTTDGLARLDIDETEHADGKVPYIKM 895
 DB 836 AKGMSYLEDVRLVHNDLAARNVLVSPNHNKTTDGLARLDIDETEHADGKVPYIKM 889
 QY 896 ALESILRRRTFTHOSDVMSYGVTVMEMLTFGAKPYGCIIPAREIPDLEKEGERLPDPICTI 955
 DB 896 ALESILRRRTFTHOSDVMSYGVTVMEMLTFGAKPYGCIIPAREIPDLEKEGERLPDPICTI 949
 QY 956 DVTMIMVCMWIDSECRPRFELVSEFSRMARDPQRFVITONEDLGSPASPLDSTYRSLL 1015
 DB 956 DVTMIMVCMWIDSECRPRFELVSEFSRMARDPQRFVITONEDLGSPASPLDSTYRSLL 1009

QY 1016 EDDMDGLVDAEELVLPQGGFCPPDAPGAGVHHRRSSSTRSGGDLTLGLPESEE 1075
DB 1010 EDDMDGLVDAEELVLPQGGFCPPDAPGAGVHHRRSSSTRSGGDLTLGLPESEE 1069
QY 1076 APRSLAPSEAGSDVFDGLGMAKGLSPLTHDPSPLQRISEDPVLPSETDGYVA 1135
DB 1070 PRPSPLAPSEAGSDVFEGLGMAKGLSPQISPRDISPLQRISEDPVLPSETDGYVA 1129
QY 1136 PLTCSPOPEYVNPDPVPPSPREGPLPAPAPAGATLERAKTLSPGKGVVMDVFAFGG 1195
DB 1130 PLACSPQPEYVNPDPVPPSPREGPLPAPAPAGATLERAKTLSPGKGVVMDVFAFGG 1189
QY 1196 AVENPEYLVPRGSGASQHP-PPALCPAFDNLVYWDQPPSRGSPPTFEETPRAENPEYL 1255
DB 1190 AVENPEYLVPRGSGASQHP-PPALCPAFDNLVYWDQPPSRGSPPTFEETPRAENPEYL 1248
QY 1256 GLDVPV 1261
DB 1249 GLDVPV 1254

RESULT 4
EGFR_HUMAN STANDARD; PRT; 1210 AA.
ID EGFR_HUMAN Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC Q00688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
protein-cytosolic kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whitte N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells";
RT Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maibhe N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGR) in ovarian cancer";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearrell R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crosley T.D., Magnuson T.R., James C.D.,
RA Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crosley T.O., Magnuson T.R.,
RA Maibhe N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krulger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Cope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whitte N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription";
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Straton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA." ;
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Marcolis B.L., Lax I., Kries R., Dombajagian M., Hongger A.M.,
 RT Howk R., Glvol D., Ulrich A., Schlesinger J.,
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor." ;
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.,
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts." ;
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.,
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor." ;
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Otake M., Nagaki F., Lax I., Schlesinger J., Kohda D.,
 RT "Disulfide bond structure of human epidermal growth factor receptor." ;
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.,
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens." ;
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -I- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -I- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -I- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -I- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -I- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X00588; CAA25240.1; -
 CC EMBL; U95089; AAB53063.1; -
 CC EMBL; U48722; AAC50802.1; -

DR EMBL; U48723; AAC50804.1; -
 DR EMBL; U48724; AAC50796.1; -
 DR EMBL; U48725; AAC50797.1; -
 DR EMBL; U48726; AAC50798.1; -

Query Match 45.5%; Score 3116; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 2,7e-159;
 Matches 630; Conservative 177; Mismatches 350; Indels 116; Gaps 25;

QY 11 LLLALLPPGAA--STVCGCTGDMKRLPASPETHLDMLRLYGGCCVGVGQVGNLETTYPTN 68
 Db LLLALLPPGAA--STVCGCTGDMKRLPASPETHLDMLRLYGGCCVGVGQVGNLETTYPTN 73
 QY 69 ASLSFLQDIQEVGVYLLAHNQVQVPLQRLIRVGTQLEFDVYVALAVLDNGPLNNTTP 128
 Db YDLSFLKTIQEVAVGVYLLAHNTYERPLENLQIIRGMYYENSYALAVLSNYD----- 126
 QY 129 VVGASPGIARELDLRLSLTEILKGVLIQRNPOLCYODTIWKDI FHKNNQALTLIDTNR 188
 Db 127 ---ANKTGKELPMRNLOEILHGAVRFSSNNPALCNVESIQWRDIVSSDFLSNMSMDPNH 183
 QY 189 SRACHSPCMCKGSRGCMGSSSEDDCSLTTRTVAGGCA-RCKGPLPDDCCEGCAAGCTG 247
 Db 184 LGSQCKQCDPSCPNGSCWGAEEGCKLTXTIKQCCSGRCRGKSPSDCCNQCAGCTG 243
 QY 248 KHSDDLACLAHPNHSIGICELCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNYLSTD 307
 Db 244 RSDDLVCRKFRDEACXCTCPPLMYNPTTYGMDVNPBGKYSFGATCYVCKCRNRYVTD 303
 QY 308 VSGCTLVCPLRHNOETVAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGITELE-FAG 366
 Db 304 HSCVAVACADSYEM-EEDGVKCKKCEPCRCVNCNGIGERK-DLSINATNIKFKN 361
 QY 367 CKKIFSLFLPSPFCGDDASNTAPLQEPOLQVFELEETETGLYISAWDSPLDSVFG 426
 Db 362 CTISGDLHILPAFEGDSFTHTPPLDPELDLTKVKETGFLLIQAMENNTDLAF 421
 QY 427 NLOVIRGRILHNGAVSLTQGLGISWIGRLSRELSSGLALIHNNHLCFVHTVPVDPQLF 486
 Db 422 NLEIRGRKQHPFLAVLSINTSLGRSLKEISDGVYIISGNRLCYANTINMKL 481
 QY 487 RNNFTVSFWLAVPKVASLHLE-ECVGEGLACHQLCARGHCWGPPTOCVNCSSQFLRGQ 545
 Db 482 GTSG-----OKTKIISRGNSCAQGVCHALCSPGCMGPRDVCSCNVSGRE 534
 QY 546 CYBECKVLOGLRREYVNAHCLPCHECCOPNGSVTCFSPBADCCVACAHYKPPCVAR 605
 Db 535 CVDKCNLEGEPRFEFENSECTIQCHECLPQAMNITCTGRGPNCITQCAHYIDGPHCVKT 594
 QY 606 CPSGVPRDLSYMPIMKFPDEEGACQPCPINCCHSCVDLDDKGPAGORASPLTSISAVY 665
 Db 595 CPAGVWGENNTL-VKRYADAGVCHLCHNCTYCTGPGLEGCTTPGPKIP--SIATGV 651
 QY 666 G---LLVVVLGVVFGILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNOQRILK 722
 Db 652 GALLLLVVALGIG--LFMRRRHIVRKTLRLLOERELVEPLTPSGAMPNOQRILK 708
 QY 723 ETELKRVKVLGSGAGTYVKGKWIPIDEGVNKKI PVAIKVLRNTSPKANKELDEAYVWAG 782
 Db 709 ETEFKKIKVLGSGAGTYVKGKWIPIDEGEVKI PVAIKELRATSPKANKELIDEAYVWAS 768
 QY 783 VGSFVYSRLGICLSTVQLVTLQMLPYGCLLHVVENRGLSODLLNMCQIAKMSVY 842
 Db 769 VDNPHVCRLLGICLSTVQLVTLQMLPYGCLLHVVENRGLSODLLNMCQIAKMSVY 828
 QY 843 EDVRLVHDLAARNVLYVSPNFKITDFGLARLDIDETRYHADGGKVPKIMMALESILR 902
 Db 829 EDRLVHDLAARNVLYVTPGHVKITDFGLAKLGAEEKEVHAEGKVPKIMMALESILH 888
 QY 903 RRFTHOSVMSYGVTVWELMTFGARPYDGIIPAREIPDLLEKGERLPQPICTIDVYMINV 962
 Db 889 RIYTHOSVMSYGVTVWELMTFGSKPYDGIIPASEISISLEKGERLPQPICTIDVYMINV 948

Qy	963	KCMIMDSCRFRELVEEFSRMADPORFVIO-NEDIGASPLDSFYFSLLEDDMG	1022
Db	949	KCMIMDASRRKFRLLIIEFSGMARDPORVLIYIOQDEKHNLEPFDNSFYRLMDEMD	1006
Qy	1022	DLVDAEYLVRQOGFCFCDPAPGAGGMVHHRRSSSTSGGDLTLGLEPSEEARPSPL	1081
Db	1009	DVVADEYLLPQOGFF-----SSPSTRSRPL	1034
Qy	1082	APSEGAGSDVDRDGLGMGAAGKLOSLPTHDSPLORSEDPVLPSEI--DGVAAPLTC	1139
Db	1035	LSLSLATSIN--NSTVACIDIRNGLOSCPIKEDSFLORYSDSDPTGALTEDSIDDTFL--	1087
Qy	1140	SPOREYVNVQRPVRRPSPRESQPLPAARPAGATLERAKTILSPGKGVVYKDNPAFGAVEN	1199
Db	1088	RPVEYVINO-SVPRKPAQSVONPVYHNPFLND-----APSRDPHYOD--PHSTAVGN	1139
Qy	1200	PEYLV-TPQSGAARQENRRPAEFPADNLVYMDQ-----DP-----PERGAPST	1244
Db	1136	PEYLVTVQ-----PTCVNSTSTDSRAHMAQKSHQISLDNPVYQODFFPKKAKRNGI	1166
Qy	1243	FKGTPTLANPEYLV	1255
Db	1187	FKGS-TAENAEYLV	1198

ID	EGFR_MOUSE	STANDARD	PRT	1210 AA.
AC	001279;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112).			
EGFR				
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/c; TISSUE=Liver;			
RC	MEDLINE=93026370; PubMed=1408137;			
RX	Avvati A., Skorecki K., Yayon A., Givol D.;			
RT	"Promoter region of the murine fibroblast growth factor receptor 2			
RT	(bek/KGFR) gene.";			
RL	Oncogene 7:1957-1962(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;			
RX	MEDLINE=93126380; PubMed=7678348;			
RX	Paria B.C., Dae S.K., Andrews G.K., Dey S.K.;			
RT	"Expression of the epidermal growth factor receptor gene is regulated			
RT	in mouse blastocysts during delayed implantation.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/c; TISSUE=Liver;			
RC	Hibbs M.L.;			
RL	Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=B6/C3; TISSUE=Liver;			
RC	MEDLINE=94170986; PubMed=6125255;			
RA	Luetkeke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,			
RA	Jenkins N.A., Lee D.C.;			
RT	"The mouse waved-2 phenotype results from a point mutation in the EGF			
RT	receptor tyrosine kinase."			
RL	Genes Dev. 8:399-413(1994).			
RN	[5]			
RP	SEQUENCE OF 1-714 FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=91232866; PubMed=2030916;			
RA	Aviv A., Lax I., Ulrich A., Schlesinger J., Givol D., Morse B.;			

```

RT "Comparison of EGF receptor sequences as a guide to study the ligand
RL binding site."
RN Oncogene 6:673-676(1991).
RP [6]
RC SEQUENCE OF 969-1117 FROM N.A.
RA STRAIN=C3H;
RA Eisinger D.P.; Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as the content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DR		EMBL; X78987;	CAA55587.1 ; -	-
DR		EMBL; U03425;	AAA17899.1 ; -	-
DR		EMBL; X59698;	CAA42219.1 ; -	-
DR		EMBL; L06864;	AAA53029.1 ; -	-
DR		EMBL; Z12608;	CAA78249.1 ; -	-
DR	HSSP ;	P11362 ;	IFGK.	
DR	MGD ;	MGI :95294 ;	Egfr.	
DR	InterPro :	IPR000494 ;	EGFR_L domain.	
DR	InterPro :	IPR000719 ;	Euk_DKinase.	
DR	InterPro :	IPR002174 ;	Furin-like.	
DR	InterPro :	IPR001245 ;	Tyr_kinase.	
DR	pfam ;	PF00069 ;	Dkinase ; 1.	
DR	pfam ;	PF00757 ;	Furin-like ; 1.	
DR	pfam ;	PF01030 ;	Recep_L domain ; 2.	
DR	ProDom ;	PD000001 ;	Euk_pkinase ; 1.	
DR	SMART ;	SMO0261 ;	FU ; 3.	
DR	SMART ;	SMO0219 ;	TyrcKc ; 1.	
DR	PROSITE ;	PS00107 ;	PROTEIN_KINASE_ATP ; 1.	
DR	PROSITE ;	PS00109 ;	PROTEIN_KINASE_TYR ; 1.	
DR	PROSITE ;	PS50011 ;	PROTEIN_KINASE_DOM ; 1.	
KV	Transmembran ;	Glycoprotein ;	Receptor ; Signal ; Transferase ;	
KW	Tyrosine-protein kinase ;	ATP-binding ;	phosphorylation ; Repeat .	
FT	SIGNAL	1	24	POTENTIAL.
FT	CHAIN	25	1210	EPIDERMAL GROWTH FACTOR RECEPTOR.
FT	DOMAIN	25	647	EXTRACELLULAR (POTENTIAL).
FT	TRANSEM	648	670	POTENTIAL.
FT	DOMAIN	671	1210	CYTOPLASMIC (POTENTIAL) .
FT	REPEAT	75	300	APPROXIMATE.
FT	REPEAT	390	600	APPROXIMATE.
FT	DOMAIN	1028	1071	SER-RICH.
FT	DOMAIN	714	981	PROTEIN KINASE.
FT	NB_BIND	720	728	ATP (BY SIMILARITY) .
FT	BINDING	747	747	ATP (BY SIMILARITY) .
FT	ACT_SITE	839	839	BY SIMILARITY.
FT	DISULFID	190	199	BY SIMILARITY.
FT	DISULFID	194	207	BY SIMILARITY.
FT	DISULFID	215	223	BY SIMILARITY.
FT	DISULFID	219	231	BY SIMILARITY.
FT	DISULFID	232	240	BY SIMILARITY.
FT	DISULFID	236	248	BY SIMILARITY.
FT	DISULFID	251	260	BY SIMILARITY.
FT	DISULFID	264	291	BY SIMILARITY.
FT	DISULFID	295	307	BY SIMILARITY.
FT	DISULFID	311	326	BY SIMILARITY.
FT	DISULFID	329	333	BY SIMILARITY.

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 991 991 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF22F5 CRC64;

```

```

Query Match 45.1%; Score 3087; DB 1; Length 1210;
Best Local Similarity 49.4%; Pred. No. 9.6e-158;
Matches 632; Conservative 170; Mismatches 358; Indels 120; Gaps 27;

```

```

QY 11 LLLALLPRGAA--STVCTGTDMLRLPASBETHLDMRLHYQCCGYVQGNLEITLPTN 68
DB 14 LLLALLCAAGALEEKVCQGSNSRLTQLGFEDHFLSLQRVYNCEVVLGLLEITVYQNR 73
QY 69 ASLSFLQDIOEVQGVYLIANNQVQVPLQRLRIVRGTQLFEDNALVAVLNDGPDLANNTTP 128
DB 74 YDLSFLKTIQEVQGVYLIANNQVQVPLQRLRIVRGTQLFEDNALVAVLNDGPDLANNTTP 124
QY 129 VTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTILMKDI----FKKNQALATLI 184
DB 125 -YGTNRGTGLRELPMRNLOEILIGAVRSPNPILCNMDTIQWRDVLQVAVFMSMSMDL--- 180
QY 185 DTRRSRAHRCSPMKCKSRCKWGESSEDCSLTRVCGGCA-RCKGRLPTDCHEGCAAG 243
DB 181 -QSPSSCPKCDSPCKNGSCWGGEEENCKLTKLICQOCSHRCGRSPSDCHNQCAAG 239
QY 244 CTGPKSDCLACLHFNHSGICELHCPALVYNTDTEFSMNPREGVTFEGASCYACPNY 303
DB 240 CTGPRSDCLVCKQFQDEATCKDTCPLMLYNTPTTYMDVNPBGKISFGATCYAKKCPRY 299
QY 304 LSTDVSCSLVCPRLHNOEVTABDGTORCEKSKRCARVCGYLGQVYKANSKFIETELE 363
DB 300 VVTDHSCSVACGCPDYEV-EEQGIKCKCKDCGCRKVCIGIGERK-DTSLINATNIK 357
QY 364 -PAGCKKIFGSLAFLESFPGDASNTAPLQPEOLQFETLEITGVLYISAMPDLPDL 422
DB 358 HFKYCTAISGDHLILPVAFKGDSPTRPPDLPRELEILTKVKEITGFLILQAMPDMWTDL 417
QY 423 SVFQNTQVIRGRILHNGAYSLTLOGIGISWLGRLSRLBESGLALIHNNHLCPVHTVPM 482
DB 418 HAFENLEIRGRKQKQGFSLAVGANTISLGRSLKEISDGVYIISGNHNLCTYANTINM 477
QY 483 DQLFNNFTVSWLVRPVKVSASHLE-ECYEGELACHOLCARGHCMGPGPTQCVNCSQFL 541
DB 478 KKLFGTPN-----OKTKIMNRAREKDKAVNHVNCPLCSSGCGPREDVCVQGNVS 530
QY 542 RQGECEBECRVLOGLREYVNAHCLPCHNECQRPQNGSVTCFGRBADQCVACAHYDDPP 601

```

```

DB 531 RGRCEVCKNLTLEGEPRFVENSECIOCHPECLPQANMITCTGKRPDNCIOCAHYIDPH 590
QY 602 CVACPSGVKDLSTYPMWKPRDEGACQPCPINCTHS CVLDKQCPAEGRASPSTIV 661
DB 591 CVKTCPPGIMENNTL-VMKYADANNVCHLCHACCTGACGPGLOGCVPWSPGPKTISA 649
QY 662 SAVGILLVVLGVVFGI-LIKRQOKIRKTYMRLLOETELVEFLTPSGAMPNOQMRI 720
DB 650 TGIYGLGLFTIV-VALGIGLFMRRRHIVKRTKTLRLLOERLVELTETSGEAPNAHLRI 708
QY 721 LKETELRKVKYLSGAGFTGVYKGIWIDGENVKIPVAIKVLENTSPRANKELIDEAYVM 780
DB 709 LKETEFKKIKYLSGAGFTGVYKGIWIDGENVKIPVAIKVLENTSPRANKELIDEAYVM 768
QY 781 AGVSPVSVRLGICLTSTVOLYQMLPYGCLLDHNEBRRLSSODLLNMCQIAKMS 840
DB 769 ASVNPVPCRLGLCTSTVOLYQMLPYGCLLDYVBRHKNISQYLLNVCQIAKMN 828
QY 841 YLEBDRVLVHRDLAARNVLYKSPNHVKITDFGLARLDIDETRYADGGKVPILKMMALESI 900
DB 829 YLEBDRVLVHRDLAARNVLYKTPQHVKITDFGLARLDIDETRYADGGKVPILKMMALESI 888
QY 901 LRRRFTQSDVMSYGVTVWELMTFGAARPYDGIIPAREIPDLLEKGERLPQPCITIDVYMI 960
DB 889 LHRITYHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPCITIDVYMI 948
QY 961 MYKCMWIDSECRPRFRRLVSEFSNMARDPQRFVYIQ-NEDGSPSPLDSTYRSLIEDD 1019
DB 949 MYKCMWIDSECRPRFRRLVSEFSNMARDPQRFVYIQDERHNLSPDSTYRSLIEDD 1008
QY 1020 MGDVLAEEYLVPOQGFPCDPAPGAGMYHHRSSSTRSGGDLTLGLPSEEARNS 1079
DB 1009 MEDVYVDELVLIPQGF-----NSST-----SRT 1034
QY 1080 PLAPSEGAGSDVPEGDLGKMAKGLQSLPTHPSPLQRYSEDPVLPSEF-DGYVAPL 1137
DB 1035 PLTSLISATSN---NSTVACINRNGSCRVKEDAFGLQRYSSDPGAVTEBIDIDAF--- 1087
QY 1138 TCSPPQRYVQNPQVPRQPPSPREGRPLPAARPAAGTLERAKTLISGKGVKQVADVAFGAV 1197
DB 1088 ---PVPEYVNO-SVPKRPAGSVQNPVYHNPRLP-----ABGRDLHYON--PHSNVAV 1133
QY 1198 ENPEYL-TPOGGAAPQHPHAPSPAFDNIYUDD-----DP-----PERGAP 1240
DB 1134 GNEBYLNTAQ-----PTCLSSGSPSALMTIQKSHQMSLDNPYQODFFPKETKP 1184
QY 1241 STFKGTPTAENPEYGLDVP 1260
DB 1185 GIFKG-PTAENAEYLRVAP 1203

```

```

RESULT 6
ERBA HUMAN STANDARD; PRT; 1308 AA.
ID ERBA HUMAN
AC 015303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase etbB-4 precursor (BC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERBA OR HER4.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Pionman G.D., Culouscou J.-M., Whitley G.S., Green J.M., Carlton G.W.,
RA Foy L., Neudauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

```

RT epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RA MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Cortas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768 (1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CREBBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PANATHROID, CREBBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND PANCREAS.
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L07868; AAB59446.1; -.
 DR HSSP: P11362; IFCK.
 DR Genew: HGNC:3432; ERBB4.
 DR MIM: 600543; -.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR Prodom: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00261; Fu; 4.
 DR SMART: SM00219; Tykic; 1.
 DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN KINASE_DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 PROTEIN KINASE.

FT NP BIND 724 732
 FT BINDING 751 751
 FT ACT SITE 843 843
 FT DISULFID 189 197
 FT DISULFID 193 205
 FT DISULFID 213 221
 FT DISULFID 217 229
 FT DISULFID 230 238
 FT DISULFID 234 246
 FT DISULFID 249 258
 FT DISULFID 262 289
 FT DISULFID 293 304
 FT DISULFID 308 323
 FT DISULFID 326 330
 FT DISULFID 503 512
 FT DISULFID 507 520
 FT DISULFID 523 532
 FT DISULFID 536 552
 FT DISULFID 555 569
 FT DISULFID 559 577
 FT DISULFID 580 589
 FT DISULFID 593 614
 FT DISULFID 617 625
 FT DISULFID 621 633
 FT MOD_RES 1162 1162
 FT MOD_RES 1188 1188
 FT MOD_RES 1258 1258
 FT MOD_RES 1284 1284
 FT CARBOHYD 138 138
 FT CARBOHYD 174 174
 FT CARBOHYD 181 181
 FT CARBOHYD 253 253
 FT CARBOHYD 358 358
 FT CARBOHYD 410 410
 FT CARBOHYD 473 473
 FT CARBOHYD 495 495
 FT CARBOHYD 548 548
 FT CARBOHYD 576 576
 FT CARBOHYD 620 620
 FT VARSPLIC 626 648
 FT SO SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;
 Query Match 43.0%; Score 2942.5; DB 1; Length 1308;
 Best local Similarity 45.0%; Pred. No. 5.5e-150;
 Matches 609; Conservative 181; Mismatches 383; Indels 179; Gaps 29;
 QY 9 WGLLIALLPQGA-----STOVCTGTDMKRLRPASPEETHLDMRLHYOCQVVOGNTLELY 64
 DB 8 MWWSLLVAAGTVPSPDSQSCAGTENTLSLSDEQYRALRKTYENCEVVMGNLEITS 67
 QY 65 LPTNASLSFLDIOEVQGYVLIANOVQVPLQRLIRIVRGTOLEEDNYALAVLDNGDPLN 124
 DB 68 IEHRDLSFLRSVRELVGYVALNQFERYLPLEMLIRIGTKLYEDRYALAIFLNVRKDG 127
 QY 125 NTPVPTGASPGCLAELOLRSTLTILKGVYLQRPOLCYOITILMKDIFHKONLALILI 184
 DB 128 NF-----GLDELGLKNTLETLNGGVYVDKFLCYADTINQDITVRNPMPSNLTIV 178
 QY 185 DYNBRACHPCSPCKSGSRGWESSEDSQSLTRFYVCAAGC-ARCKGPLPTCCCHQCAAG 243
 DB 179 STNSSSGCGRHKSCGTG-RCKGPEHNKQTLIRTVCAEQCGRCYGPVSDCHRECAAG 237
 QY 244 CTGPHSDDCLAGLPHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCVYACPNY 303
 DB 238 CSGPKDIDCFACNMFNSGACVTCQPFYVYNPTTFLEHNFNKAITYGACVKKCPNPF 297
 QY 304 LSTVSGCTLVCLPLNDEVTAEDETQRCCKSKCAVCYGLGMQYIKANSKFIGITBLE 363
 DB 298 V-VDSSCVPRACPSKMEV-BENIGIKMKCTDTCPRACDGIIGSGLSMAQTVSSNIDK 355
 QY 364 FAGCKKIFGSLAFIPESFDDPASNTAPLOEQLOVEFTLEIRTYGYIYSAMPDPLPS 423

```

Db 356 FINCTKINGNLIFLVTHGHDPPYNAIEAIDPEKLVNFTVREITGTEGLNIQSHPNNTPDS 415
Qy 424 VFONLOVIRGRILHNGAVSLTLOGLISWLGRLSRELSGGLALIHNNHLCFVHTVPM 483
Db 416 VFSNLVTIGGRVLYSGLSLILKQGITSLQFOSLKEISAGNIYIINDNLCYHNT 475
Qy 484 QLFRRNNTVSWLAVPKVKSASHLEECVGEGLACHOLCARHCWGPFPQCVNCSQFLRG 543
Db 476 TLFSTIN-----ORIVIRDNKRAENCTAEKGVCHNLCSDDCWGPGPQCSLCSFRFSRG 529
Qy 544 QECVECEVLQGLPREPVYNAHCHLPCHPGCP-ONGSVYCFPEADQCAACHYXDPGPC 602
Db 530 RICHSCNLVDFEFREFENGSTICVBCDPQCEKMEDELTCGHGFPNCTKCSHFQKGPNC 589
Qy 603 VARCPGVKPDLSYMPWKFPEDEGACOPCPINCTHSCVDLDDKGC-----PAE 651
Db 590 VEKCPDGLQGANFS--IFKYADPRDECHCHNCCTGCGCPNTHDCTIYYPWGHSTLPH 647
Qy 652 QASPLTISVAVV-GILLVVVLGVVFGILLIRROOKIRKRYTMRLLOETELVEPLTPSG 710
Db 648 AR-TPL--IAAGVIGLFLVIVGLFAVYVRRKSIIK-KRALRLRL-ETELVEPLTPSG 702
Qy 711 AMPNAGMRILKETELRKVKYVLSGAFGVYKGIWIPDEGNKIPVATVVAENSPKAN 770
Db 703 TAPNAGMRILKETELRKVKYVLSGAFGVYKGIWIPDEGNKIPVATVVAENSPKAN 762
Qy 771 KEILDEAVYMGVGSPPVYRLLIGICTSTVQLVTOLMPYGCULLDHVRENKRLGSDLLN 830
Db 763 VEFMEDEALIMASMDPHVRLVRLGVCISPTIQLVTOLMPHGCILLEYHEKNDIISGOLLN 822
Qy 831 WCMQIAKMSYIEDVRLVHRDLAANNVLYKSPNNHYKITDGLARLLIDETEVHADQRY 890
Db 823 WCVQIAKMSYIEERLVRDLAANNVLYKSPNNHYKITDGLARLLIDETEVHADQRY 882
Qy 891 PIKMALESLIBRRFTHOSDWSYGVWELMTFCAKPYDGI-PABEIPDLKRGRLP 950
Db 883 PIKMALESLIBRRFTHOSDWSYGVWELMTFCAKPYDGI-PABEIPDLKRGRLP 942
Qy 951 PICTIDVYIMVYKCMWIDSECRPRELEVSEFSRMAPDQRFVNIQNE-LGPASPLDST 1009
Db 943 PICTIDVYIMVYKCMWIDSECRPRELEVSEFSRMAPDQRFVNIQNE-LGPASPLDST 1002
Qy 1010 FYRSLIEDDDMDGLVDAEYVLPQCGFPDPAAGCVNTHRRSSSTRSGGDLTGL 1069
Db 1003 FFONLDEEDLEDMDAEYVLP-QAFNIPRP-----IYTSRAIDSNRS-----EIGH 1050
Qy 1070 EPSEEARPS-----PLAP-SRGASGVVDGDLGMA 1100
Db 1051 SPPEAYTMSGNQFYRDGGAEGGVSVYPAFTSTIPEAVVACQATLEIFDSCCNGT 1110
Qy 1101 AKGLQSLPTHDPSPLQRYSEDPVLP-----ETDGYVAPLTCSPQEVVNOPDVAP 1153
Db 1111 LRKPAVHVQEDSSQRYSDPTVAPERSPRGELDEBYMTPMDKPROEVLNVE--- 1167
Qy 1154 QPSPREPLPAARAGATLEBAKTLSPKGVKQVAFGAVENPEYVLPQGGAAAPR 1213
Db 1168 -----ENPFVFSRR-----KNGDLQ-----ALDNPEYHNASNG----- 1194
Qy 1214 HPPPA-----FSPAPNLYVMODPPERGA- 1238
Db 1195 -PPKAEDYVNEPLVNTFANTLGAELYKNNIISMEKAKKAFDNPDMNSLPRSTL 1253
Qy 1239 -PPSTFKGTP-----AENPEYL 1255
Db 1254 QHPDYLOEYSTKYFYKQNGRIRPIVAENPEYL 1285

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulin promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN (2)
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN (3)
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulin and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
RN (4)
RP FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
RN 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
RN NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
RN NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -SUBCELLULAR LOCATION: Type I membrane protein.
CC -TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF041838; AAD08899.1; -
DR EMBL; U52531; AAC53051.1; -
DR HSSP; P11362; IFCG.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyf_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.

```

DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TYRKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; signal;
KW tyrosinase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT DOMAIN 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
FT NP_BIND 724 732
FT BINDING 751 751
FT ACT_SITE 843 843
FT DISULFID 189 197
FT DISULFID 193 205
FT DISULFID 213 221
FT DISULFID 217 229
FT DISULFID 230 238
FT DISULFID 234 246
FT DISULFID 249 258
FT DISULFID 262 289
FT DISULFID 293 304
FT DISULFID 308 323
FT DISULFID 326 330
FT DISULFID 503 512
FT DISULFID 507 520
FT DISULFID 523 532
FT DISULFID 536 552
FT DISULFID 555 569
FT DISULFID 559 577
FT DISULFID 580 589
FT DISULFID 593 614
FT DISULFID 617 625
FT DISULFID 621 633
FT MOD_RES 1162 1162
FT MOD_RES 1188 1188
FT MOD_RES 1258 1258
FT MOD_RES 1284 1284
FT CARBOHYD 138 138
FT CARBOHYD 174 174
FT CARBOHYD 253 253
FT CARBOHYD 358 358
FT CARBOHYD 410 410
FT CARBOHYD 473 473
FT CARBOHYD 495 495
FT CARBOHYD 548 548
FT CARBOHYD 576 576
FT CARBOHYD 620 620
FT CONFLICT 1062 1062
FT CONFLICT 1080 1082
SQ SEQUENCE 1308 AA; 146957 MW; D944BB096A08B41 CRC64;
Query Match 42.7%; Score 2924; DB 1; Length 1308;
Best Local Similarity 44.9%; Pred. No. 5,4e-149;
Matches 607; Conservative 188; Mismatches 392; Indels 166; Gaps 29;
Db 1 MEIA-ALCRWGLL--ALPPGAASOVCTGTDMKRLRLPASPETHLDMLRHLYOGGOWO 57
1 MRLATGLWVGSLVLAARTVQPSASQSVACGTENKSLSLDEQVRLARKYENGEVVM 60
Qy 58 GNLELTLPFTNASLSTLQDIQEVGYVLAHNOVROVPLQRLRIIVGTQLFEDNVALVL 117
Db 61 GNLEITSIHNRDLSTLRISIREVTGYVALNQFVLPLENLRIIRGKLYEDRVYALAF 120
Qy 118 DNGDDPLNNTTPTVGAAPGGLRELOASTLEILKGGVLIQNPOLCVQDITIMQDTHKAN 177
Db 121 LMYRKDGNF-----GLQELGLKQLTEILNGGVYVDONKFLCYADTTHMODIVRNPM 171

Qy 178 QLALTLIDTRSRBACHPCSPMKSGSRCKSGESSSEDCOSLTRVCAGGC-ARCKGRLPTDC 236
Db 172 PSMNTLVSTIGSSGCGRCHSCTG-RCMGTENHCQTLLTTCVACQDQGCYGPVYSDCC 230
Qy 237 HEQCAAGCTGPKRSDCLAFHNSGICELHCPALVTYNTDTEFSMNPREGRYTFGASCV 296
Db 231 HRCACGAGCGGCKTDFACACNFDSGACVTCQCFOTFVYNTPTTQLEHNFAKTYGAFV 290
Qy 297 TACPVYLTSDVSCVLVCPYLAHQEYTAEGTORCEKSGPCARVCYGLCMQYIKANSKF 356
Db 291 KCPENHVV-VDSSCRACRCPSSXMEV-EENGIMCKRCPTDIPCACDGTGSLMSQTV 348
Qy 357 IGITTEFAFGCKKIFGSLAFLPESFDGDDPANSNAPLOPELOVETLEITGYLYICAMP 416
Db 349 DSNINDKFINGCTINGNLIPLVYTGIDPNAIDALDEPKANVFRYREITGFINTQTP 408
Qy 417 DSLPDLVSFQNLQVIRGRILHNGAYSVLQGLGISWGLSLRSLGSLALIHNTILCF 476
Db 409 PNMPTDSVFNLTIGRVLVYSGLSLLIKQGGITSLQFSLKEISAGNIYITDINSULCY 468
Qy 477 VHTVPDQLFRPNFTVSVFRLVPKVSASHLEECVGBGLACHOLCARHGCWGPPTCCVN 536
Db 469 YHTINNTTLFTSTVN-----QRIVIRDNRAENCTAEGMVCNHLCSNDGCMGPQDCLS 522
Qy 537 CSQFLRGQECVEECRVLQGLPREYVABHGLRCHRECP-ONGSVTCFGEADQVACAH 595
Db 523 CRFSSGKCTIESCNLYDGEFRFENGSLICEVDSQCEKMDGLTCHGGRDCTKSH 582
Qy 596 YKDPRECVARCPGVPDLVSYMPIMKFPDEGACORPINCSTHSCVDLDDKGC----- 648
Db 583 FKQGRPCVEKCPVLQGANF--IFKADDRCHRPCHRCCTGCGNPTSHDCIYRPMTG 640
Qy 649 ----PAPQASPLTSTVSANV-GILVVVLGVVFGILYKROOKIRKTYMRLLQETELV 703
Db 641 HSTLPQPHAR-TPL-IAAGVIGLFLVLMALTFVAVVRRKSIK-KKRALRFL-ETELV 695
Qy 704 EPLTPSGAMPNOQMILKETELRKVYGVGSGAGFQTYKGIWIPDGENVYIPVAIKYURE 763
Db 696 EPLTPSGTAPNOQLILKETELRKVYGVGSGAGFQTYKGIWIPDGENVYIPVAIKYURE 755
Qy 764 NTPSKANKELDEAYVAGVSPVSRLLGICLTSTVOLVTOIMPYGLLDHYENRGRL 823
Db 756 TTPCRKANVEFMDALIMASVDHPLVRLIGVCSPTIQLVTOIMPYGLLDHYENRGRL 815
Qy 824 GSDDLANKCQIAKGSYLEDVLYNRDLAARVLYKSPNHVKITDFGLARLLDIDETEV 883
Db 816 GSQLLNMCVQIAKGMVYLEERLYVNRDLAARVLYKSPNHVKITDFGLARLLDIDETEV 875
Qy 884 HADGKVPIKMALESILRRPFTHSDVMSYGVTVLMLTFGAKPYDGIPIAREIPDLLEK 943
Db 876 NADGKMPKIMALECIIHRTKFTHSDVMSYGVTVLMLTFGAKPYDGIPIAREIPDLLEK 935
Qy 944 GERLPQPICTIDVYIMVYKMWIDSECPREFRELYSEFSRMAADPQRFVIONED-LGP 1002
Db 936 GERLPQPICTIDVYIMVYKMWIDSECPREFRELYSEFSRMAADPQRFVIONED-LGP 995
Qy 1003 ASPLDSTFRLSLDEDDMDGLVDAEYLYVPQGFCCPDP----- 1041
Db 996 PSFNDKSFQNLDEDEDLDDMDDAEYLYVP-QAFNIPPIYRTSRTRIDSNRSEIGHSPRP 1054
Qy 1042 --APGAGGVNHHNRHSSRSRSGGDLTLGLSEBEERAPRPLAPSGAGSDVVDGDLGMC 1099
Db 1055 ATTPMSSGQFVVDGDFATQCG--NPMRYTATTSITPEAPVA--QGATLMEFDDSCNG 1109
Qy 1100 AAKGLQSLPTHPSPLOQRYSEDPVLPSP-----ETDGYVPLTCSPOPEYVNOQDVR 1152
Db 1110 TLKRPVPHVVOEDSSQRYASAPPTVAFARNPRPALEDESGYTPMMDKPKQVETLNPVE-- 1167
Qy 1153 POPSPREGRPLPAARPAQATLEBAKTLSPKNGSVKDVAFPGAVENPEVLTPOGGAAPQ 1212
Db 1168 -----ENPFPVSR-----KNGDLO-----ALDNPENHASGSG----- 1194
Qy 1213 PHPPRA-----FSAPFQNLVYWDPPPERGA 1238

```

Db      1195 --PPKADEYVNEBLYLNTFTNALGNAYMKNSLSVBEKAKKAFDNDPYNHSLPFRST 1252
Oy      1239 --PSTFGKPT-----AENPEYL 1255
Db      1253 LQHPDYUQEVSTKYFKONGRIKRPVLENPEYL 1285

RESULT 8
XMRK_XIPMA STANDARD; PRT; 1167 AA.
ID      XMRK_XIPMA
AC      P13388
DT      01-JAN-1990 (Rel. 13, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN      XMRK OR TY.
OS      Xiphophorus maculatus (Southern platyfish).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC      Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC      Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX      NCBI_TaxID=8083;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90015140; PubMed=2797166;
RA      Witbrodt J., Adam D., Malitschek B., Maueiler W., Raulf F.,
RA      Telling A., Robertson S.M., Schartl M.;
RT      "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT      inducing Tu locus in Xiphophorus.";
RL      Nature 341:415-421 (1998).
RN      [2]
RP      REVISION TO 515.
RA      Schartl M.;
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC      -1: FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC      -1: CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1: SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1: DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC      -1: SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; X16891; CAA34770.2; -
DR      PIR; S06142; S06142.
DR      HSSP; P13362; IRGK.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_Pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR002290; Ser_thr_kinase.
DR      InterPro; IPR001245; Tyr_kinase.
DR      Pfam; PF00069; Pkinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00261; FU; 5.
DR      SMART; SM00220; S_TKC; 1.
DR      SMART; SM00219; TYKC; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM      Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM      Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT      SIGNAL 1 25
FT      CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT      DOMAIN 26 642 KINASE.
FT      TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
FT      DOMAIN 666 1167 POTENTIAL.
FT      DOMAIN 1167 1285 CYTOPLASMIC (POTENTIAL).
FT      BIND 710 977 PROTEIN KINASE.
FT      BIND 716 724 ATP (BY SIMILARITY).
FT      BIND 743 743 ATP (BY SIMILARITY).
FT      ACT_SITE 835 835 BY SIMILARITY.
FT      DISULFID 195 204 BY SIMILARITY.
FT      DISULFID 199 212 BY SIMILARITY.
FT      DISULFID 220 228 BY SIMILARITY.
FT      DISULFID 224 236 BY SIMILARITY.
FT      DISULFID 237 245 BY SIMILARITY.
FT      DISULFID 241 253 BY SIMILARITY.
FT      DISULFID 256 265 BY SIMILARITY.
FT      DISULFID 269 296 BY SIMILARITY.
FT      DISULFID 300 311 BY SIMILARITY.
FT      DISULFID 315 330 BY SIMILARITY.
FT      DISULFID 333 337 BY SIMILARITY.
FT      DISULFID 504 513 BY SIMILARITY.
FT      DISULFID 508 521 BY SIMILARITY.
FT      DISULFID 524 533 BY SIMILARITY.
FT      DISULFID 537 553 BY SIMILARITY.
FT      DISULFID 556 569 BY SIMILARITY.
FT      DISULFID 560 577 BY SIMILARITY.
FT      DISULFID 593 615 BY SIMILARITY.
FT      DISULFID 618 626 BY SIMILARITY.
FT      DISULFID 622 634 BY SIMILARITY.
FT      CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 39.1%; Score 2677.5; DB 1; Length 1167;
Best Local Similarity 44.9%; Pred. No. 7,1e-136;
Matches 572; Conservative 167; Mismatches 390; Indels 145; Gaps 29;

Oy      4 AALCRWGLLALLPFGAAS-----QVCTGTDMKRLRPSPEHLDMLHLVGGCOVQGN 59
Db      8 AALLQ--LLVLVLSIRCCSTDPDRKVCQGTSNQMTM---LDNRYLYMKRMVSGCNVAVLEN 62
Oy      60 LELTYLPTNASTLSEFLODIOEVGYVLIHNOVQVPLQRLRIVRGTOLFEDNYALAVLDN 119
Db      63 LEITYQENQDLSFLQSIQEVGYVLIHNEVSTIPLVNLRIKGNLYEGNFTLLVMSN 122
Oy      120 GDPINNTTPVTGASPGGIRELQSLRLTEILKGVLIOHNPOLCYODTLLMKDIFKHQNL 179
Db      123 YQK-NPSSP--DYQVQGLKQQLSVLTETILSGVYVSHNPPLLCVETINMMDIVDKTSNP 179
Oy      180 ALTLIDTRRSRAHCPSCMCKGSKCGSSSEDQSLTTVCAAGG-AACKGPLPDDCHE 238
Db      180 TWNLIPHAFERQCKCKCHGCVNGSCWAPGHCCKFTLLCAEQCNRCRCRGPIDCCNE 239
Oy      239 QCAAGCTGPKSKDGLACLFHNSGICELCPALVYTNDDTESMNPGRGTFGASCYTA 298
Db      240 HCAGGCTGPRATDLCACDFDDGTCKQTCPPKITYDIVSHQVVDNPIKITTFGAACVYE 299
Oy      299 CPYNYLSTDVGSCTLVCEPLHNOEYTABDGTORCEKSKPCARVCYGLGMOYIKANSKEIG 358
Db      300 CPSNYVYTE-CACVRSASAGMLEVD-ENGKRSCKPCDQVCPKVCIGDIGSL-SNTIAVN 356
Oy      359 ITTEL-EPAGCKKIKGSLAPLBPESDGPASNTALPQLOVFTLEITGVLISAMPD 417
Db      357 STNIRSFNCTCKKINGDIILNENSFEGDPHYKIGTMDPEHLNLTLYVEITLYVLWMPPE 416
Oy      418 SLPLSVFQNLQVIRGRILHNHGAYS-LTLQGLGISWLGRLREISGLALIHNNTHLCF 476

```

```

Db      417  NMTSLVFPQLEIIIGRTTFSRGFSFVVVQVRLQWLMGLRSLSKEYSAGNVILKNTLQLRLY 476
      477  VHTVPMWDLFRFNNNTVTSFWLVRPKVASHLBECVGEGLACGLQCARGHGCMRPTQCVN 536
      477  ANTIIMWRLLFSESDSIEYDART-----ENQTCNNESSEDCMWPGRPTMCS 523
      537  CSQFLRGOECVEECVNLQGLPREYVNAHCLPCHPECCOPQNSVTCFGEADQCAVCAHY 596
      524  CLHVDGRGCVAASCHLQGEPRBAQVDGRCVQCHQECVLTQVTSILCTYGFGRNCSKSAHF 583
      597  KDPEFCVARSQGVKPRDLSYPMWKFPBEGACORPCINCTHSCVLDLKDCKGPABORASP 656
      584  ODGPQICIRPCHGILGDDTL-IMKYADKMGCCQCHQCTGGCGSPGLSGCRGD-IVSH 641
      657  LSTVSAVGLLVVVLGVVRGILIKRROQKIRKXTMRRLLOETELVEPLTSSGAMPNO 716
      642  SSLAAGLVSGLLITVALLVLLVLRRLRRIR-KKTTICLLQEKLEVEPLTSSGAPNOA 700
      717  QMRILKETELRKVKVLSGAGFQVYKGIWIPDGENVKIPVAIKVIRENTSPANKELIDE 776
      701  FLRIKETEFKDRVLSGAGFQVYKGIWIPDGENVKIPVAIKVIRENTSPANKELIDE 760
      777  AYVWAGVSPYVSRLLGICLTSTVOLVQLMPYGLLDHVRNRRGLSQDLLNWCQOJA 836
      761  AYVWASVDPHPRVCRLLGICLTSAVOLVQLMPYGLLDHVRNRRGLSQDLLNWCQOJA 820
      837  KGMSTLEDPRLVHRDLARNVLYKSPNHVKTIDPCLALLDIDETEHADGKRVIKMA 896
      821  KGMVLEERHLVHRDLARNVLYKSPNHVKTIDPCLALLDIDETEHADGKRVIKMA 880
      897  LESILRRRTFQSDVWYGVYVWELMTFGAKPYDGIIPAREIPDLKEGERLPQPICTID 956
      881  LESILQWYTHQSDVWYGVYVWELMTFGSKYDIDIPAKELASVLENGERLPQPICTIE 940
      957  VYMWKCMWIDSECRPFRELVSEFSRWARDPQRFVVIQNEIDLPASPLDSTFYRSLLE 1016
      941  VYMIILKCMWIDPSSRPFRELVSEFSRWARDPQRFVVIQNEIDLPASPLDSTFYRSLLS 997
      1017  DDDMGDLVDAREYLVPOQGFPCRPAPAGCGVNHHRHSSSTRGCGGLTLGLESEBEA 1076
      998  SDD--DVVDADDEYLYPKRI-----NRQSS----- 1020
      1077  PRSLPASEGAGSDVFDGLMGAKAGLQSLPTHDPSLQRYSEDPV-PIPSETDGYVA 1135
      1021  --EFCIPETG-----PVRENSTLTKRNISPTQNALEKLDG-- 1056
      1136  PLTCSPOPEYVNOQVPRQP-----PSPRE--GPLP-AARPAATLERAKTSLSPG 1182
      1057  -----EYVNOQSETSRSLSDIYNPNYEDLTDGMGPVSLSQEARTFSRPEYLTNTN 1108
      1183  KNGVVKDVFAGAVENPEYLTPOGGAAPQHPHPAPFAPDNLVYVQDDPERGAPST 1242
      1109  QNSL---PLVSSGSDDDPY---QAG-----YQAAF-----LPQTGALTGN 1143
      1243  FKGTPTAENPEYLG 1256
      1144  GMFLPAENLEYLG 1157

```

```

RESULT 9
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Iseling W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Toldano G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.,
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erb3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC - ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC - DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; JRGK.
DR Genew; HGNC:3431; ERB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; BgK_Dkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Dkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; Fu; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

```


KM Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 1342
 FT DOMAIN 20 643
 FT TRAMEM 644 664
 FT DOMAIN 665 1342
 FT DOMAIN 709 966
 FT NP BIND 715 723
 FT BINDING 742 742
 FT ACT SITE 834 834
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 280 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 552 565
 FT DISULFID 556 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT VARSPLIC 141 183
 FT
 FT VARSPLIC 184 1342
 FT CONFLICT 560 560
 FT CONFLICT 1064 1064
 FT SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
 Query Match 34.9%; Score 2390.5; DB 1; Length 1342;
 Best Local Similarity 40.0%; Pred. No. 1.8e-120;
 Matches 530; Conservative 196; Mismatches 448; Indels 151; Gaps 36;
 QY 10 GLLALLPRGAA--STOVCCTGDMTLRLPASPETHLDMLRLHYGCGGVVQGNLELTLYPT 67
 DB 11 GLLSLARGSEVNSQAVCPGLNGLSVTGAENQYQTLTKYERCEVVMGMLTILVTCGH 70
 QY 68 NASISFLDIOEOVGYYLIAHNOVRQVPLQRLRIYRGTLPEEDNALVALVLDNGDPLNNTT 127
 DB 71 NADISFLQWIREVGYVLANNEFSTLPLRLRVKRGVYVDGKALITVM-----LNNNT 125
 QY 128 PVTGASRGGLRELOLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNQIALLTILDTN 187
 DB 126 ----NSSHALRQLRLTQLTEILSGVYIEKNDKLCMDITMDRIDVRDRD--AETVKD 178
 QY 188 RSRACHPCSPMKCSRCMGSESDQSLTRIVYAGGC-ARKCGPLPTCCHEQCAAGCTG 246
 DB 179 NGRSCPPEHEVCCKG-RCMGPESEDCQYTLKTIICAQCNCHCFGPRPNQCCCHDECAAGCGSG 237
 QY 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPREGRYTFGASCYTACPYNYST 306
 DB 238 PPDIDCFACRHFNSGACVPRCPQPLVYNKLTJFQLEPRPHTKYQYGGVASCPRNFV-V 296

QY 307 DWGSCITLVCPRLHNOETAEEDGTORCEKSKPCARVCYGLQMOKYIKANSKE--IGITELE- 363
 DB 297 DQSSCVRACPRDMGEVD-KNGLMKCEPCGGLCPKACBEGT-----SGSRQYDSSNIDG 350
 QY 364 PACCKKIPGSLAFPLPSFGQDPASNTAPLOPEOLOVPELEBITGYLYISAMPDLS 423
 DB 351 FVNCSTKIDNLFLLTGLNGDPMHKIPALDPKTLNFRTRVEITGYLNIQSWPHMNF 410
 QY 424 VQONLOVIGRIHLNCAVS-LTLQGLGISMIGRLSRLSGSLALHNNHNLGCVTVPM 482
 DB 411 VFSNLTITGRLSYNNGFSLIMKNLNVSLGFSRLKEISAGIYISANQLCYHSLNW 470
 QY 483 DQLFRENNFTVSFWLRVP---KVSASH---LEECVEGLACHOLCARGHCGMGPTQCN 536
 DB 471 TKV-----LRGPTERRLDITGNRPDRDCAVEKVCDDPLCSSGGCGMGPGQCLS 519
 QY 537 CSQPLRGQCEVBCRYLOGLPREYVNAHRLCPHREPCOPQNSVTCFGPADQVACAHY 596
 DB 520 CKNYSRGVGVTHCNFLNGEPREFAEAEACFCSCHPFCOPMEGTATCNGSGSDTCAQAHF 579
 QY 597 KDPPECVARCPGVPKDLSTYMPIMKRPDEGACQPCPICTHSQVLDKGCFAEORA-- 654
 DB 580 RQGPFCVSSCPHGVLG--AKGPIYKYPDVQNECRPCHECTQCKGPELODCLGQTLVLI 637
 QY 655 --SPLTSIVSAVVGILLVVVLGVFGILLKRRQOKIR-KYTMRLLOETELVPLPFGA 711
 DB 638 GKTHLTMLTAVAG--LVVIPLMLGTFLYMGRRIQNKRAKRRYLERGESIEPLDS-E 694
 QY 712 MNQAOMLIKETELRKVKVLGSGAGFYVYKGIWIPDGENVKIPVAIKYLRNTSPKANK 771
 DB 695 KANKYLIARIFKETELRKLKVLGSGVGVTHKGWIPEGESIKIPVCIKYIEDSGQSFO 754
 QY 772 ELDEAYVWAGSVYVSLGICLTSTVQLYTQMLPYGCLLDHRENGRLGSDLLNW 831
 DB 755 ATDPMALAGSDHAIYVLLGLCPSSQLQVLYVLLPLSLDHDVHQHGALEPOLLNW 814
 QY 832 CMQIAKGSYLEDVRLVHRDLAARVLYVSPHNVKITDGLRLALDIDETEVHADGKVP 891
 DB 815 GVQIAKGYVLEHGMVHRNLAAKRVLLKSPQOVYADGVADLPPDDKOLLYSKATP 874
 QY 892 IKMALLESILRRRFTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
 DB 875 IKMALLESILHFKYTHQSDVMSYGVYVWELMTFGAEPYGLRLAEVPLDLLEKGERLAQ 934
 QY 952 ICTIVWMIWYCMHIDSECRFPRELYSESRMADPQRFVYIQNEDLGA--SPLDS 1008
 DB 935 ICTIDVWVWYCMHIDENIRPTFKELANEFTRMADPRLYLVIKRES-GPGIAPGPEPH 993
 QY 1009 TFRSLLEDDMDGLVDABEYLVVPOGGFPCDPAPGAGGMVHHRSSSTRGCGDITLG 1068
 DB 994 GLTNKGLAEVLEPELDLDLDEABED-----NLATTLTGSLSLP 1034
 QY 1069 LEP-SEEBAPRPLAPSEGAGSDVFDGLMGAKGLSLPTND-PSPLQRYSEDPVPL 1126
 DB 1035 VQTLNRPBGSGSLSPSSGY-MPMNQNLGSSCGSAVSGSERCRPVSILH-----PM 1087
 QY 1127 P-----SETDGYVA-----PLTSCPOPE---YVQPDVVRPQPSPRE 1160
 DB 1088 PRGLCLASSSSRGVYLSAEALQEKVSMCRSSRSRSPRPRDSAYHSQHSILTVTPLS 1147
 QY 1161 GP-----LPAARPAQATERAKTLP-SGNGV-----KDVAPFAGAVENPEY 1202
 DB 1148 PRGLEBEVNGVWPDITLKGTPSSREGTSLVGLSVLGTEEBEED-----EEYEV 1199
 QY 1203 LTPQGAAPQHPAPAFSPARDNLTYND-----ODPRPGAPPTSTFKPTPAE 1250
 DB 1200 MARRRRHSP-BHPRPSSLBELEGVRYMDVGSLSASLSTQSCPLHPRPIMTACTTDE 1258
 QY 1251 NPEYL 1255
 DB 1259 DYETM 1263

RESULT 10
ERB3_RAT STANDARD; PRT; 1339 AA.
ID_062759; 062955; 37, Created)
AC_15-DEC-1998 (Rel. 41, Last sequence update)
DT_15-JUN-2002 (Rel. 41, Last annotation update)
DT_15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein";
RL Gene 165:279-284(1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9010624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neurotrophins and their putative receptors, Erb2 and
RT Erb3, is induced during Wallerian degeneration";
RL J. Neurosci. 17:1642-1659(1997).
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U29339; AAC28498.2; -;
DR EMBL; U52530; AAC53050.1; -;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP BIND 713 721
FT BINDING 740 740
FT ACT SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 589 610
FT DISULFID 621 629
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT CONFLICT 1028 1028
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;

Query Match 34.0%; Score 2323.5; DB 1; Length 1339;
Best Local Similarity 40.4%; Pred. No. 6.9e-117;
Matches 523; Conservative 171; Mismatches 427; Indels 175; Gaps 36;

QY 3 LAALCRWGLLALLPPGAA---STQVCTGTDKKLRLPAPERTHLDMLRHLVGGCCGVVQGN 59
DB 7 LOVLC-----FLSLARGSEMGNQAVCPETTLNGLSVTGDADNOYQTLVLYKCEVWGN 62
QY 60 LELTYLPTNASLSPLODIOEVGGVYLIANOVROVPLRLRIVRGTLFEDNYALAVLDN 119
DB 63 LEIVLTGHNADLSFLQWIREVTGYLVANRERSVLPRLRVRRGVYDGRFALPVM-- 120
QY 120 GDP LANTTPVTCASRGGLRELQSLTELKGGVLIQRNPOLCYQDTILMKQI FHKNNQL 179
DB 121 ---LNYNT---NSSHALRQLKFTQLTSLSGVYIEKDKICHMDTIDMRIVRVR--- 170
QY 180 ALLTIDTRSRACHCSPMKGSRGCMGSSSSQSLTRIVVAGGC-ARCKGLPTDCCHE 238
DB 171 GAEIVYVNRNGANCPCPCHEVCKG-RKMGSGPPDCCQLIKITICAPQCGNRGFGNPNQCHD 229
QY 239 QCAAGCTGPKASDCLACLFNHSIGICELHCPALVTYNTDTFSMNPBGRVTFGASCYTA 298
DB 230 ECAGGCGSGPDPTDCPCACRRFNDGACVPRCPFLVYNTLTFQLEBNPHTKYVGGCVAS 289
QY 299 CPYNTLSDVGSCTIVCEPLHNOEYTAEDGTQRCCKSCSPCARVCTGL--GMQYIANSRF 356
DB 290 CPNHFV-VDQTFVCVARCPDRIKEDV-KHGLKRCPCGGICPRACGCTGSGSGSYQVDSN 347
QY 357 IGITELFAGCKKIFGSLAFLPESFDGDPASNTATLQPEQQLQVFETLEITTYLVISAMP 416

```

Db      348 ID-----GFWNCTKILGNDLFDLITGLANDPMHKIPALDEKLNVRTRVEITGYLNIQSWP 403
Qy      417 DSLPDLVSFONLQVIRGRILNNGAYS-LTLOGLGISWLGSLRELGLALIHNNHILC 475
Db      404 PHMNFVSFSLTTIGRSLVNRGFSLLIMKNLWTSIGFSSLSKISAGRYIASNOQLC 463
Qy      476 FVHTVPDQLFRFNNFVSWLRVPKVSASHLE-----ECVGEGLACHOLCARGHNGCP 529
Db      464 YHSHLNTRL-----LRGPSEERLDIKYDRPLGECLAGKVCDPCLSSGGCGMP 512
Qy      530 GPTQCVNCSOFLRGOECVEBCRVLOGLPREYVNAHCLCPHECOPONGSTYTCGPRENDQ 589
Db      513 GPQGLCSKRNYSRGCVVTHNCFLOGEPRFVHAQCSCHPECLMEGTSTCGSGSDA 572
Qy      590 CVACAHYKDPFPCVAVCPSPGVKPDLSYMPKPFDEEGACOPCEINTHSC--VDLDDKG 647
Db      573 CACCAHFRDGPCHCNCSPHGILG--AKGPIKYKPDQNECGRPHENCTGCGNGELDCL 630
Qy      648 CPAEQRASPLTSIVSAVVGILLVVLGVVFGILIKRROQKIR-KYMRRLLOETELVEPL 706
Db      631 GOAEVLMSKPHLVIAVVG--LAVILMILGSSFLYMGGRRIQNKRAMRYLERGESIEPL 688
Qy      707 TPGGAMPNOAMRLKETELRKVAVLGSGANGTYKGIWIPDGENVYIPIVAKYLRENTS 766
Db      689 DPS-EKANKVLARIKFKETELRKVLGSGVGVHKGIMIPESGSIKIPVCIKVIDKSG 747
Qy      767 PKANKELIDEAYVAVGSPVSRILGICLTSTVQLVTLMPYCCLDLHVENRGLRSGQ 826
Db      748 RSQFQAVTDMHVLAVGSLDHAIYRLGLCPSSIQVLTQYPLGSLDHVQHEHTLOPQ 807
Qy      827 DLNMCQIAKMSYLEDVRLVHRDLARVNLVSPNHVKITDFGLARLLDIDETEHAD 886
Db      808 LLINMGQIAKGMVYLEESHVNRDLALRNMLKSPQOVADGVADLLPPDDKQLHS 867
Qy      887 GKVPIKMMALLESILRRFTHOSDVMGSGVYVMEKMTFGAKPYGICARERIPDLLEGER 946
Db      868 EAKTPIKMMALLESILHFGKITHQSDVMGSGVYVMEKMTFGAEPYVGLAEIPDLLEGER 927
Qy      947 LPQPPICITIDVYIMVCMWIDSECRPRFELVSEFSRMAADPQRFVYIQNEDLGPA SPL 1006
Db      928 LAQPOCTIDVYVWVCMWIDENIRPTFKELANEFIRMAADPRIVIVIKRAS-GPGTP- 985
Qy      1007 DSFFYSLEDDMDGLVDABEYIVPOQGFCDPAPGAGMHHRRSSSTRSGGDLT 1066
Db      986 -PAEPSVLTTEL-----QEALELEPEL-----DLDD 1010
Qy      1067 LGLEPSE-----EAPRSPAPSE-----AGSDVFDGDLGM 1098
Db      1011 LDLEAEELATLSGLASLPTGLTLTRPGSQSLSPSSGMPMNGSLGEACLDASVLG 1070
Qy      1099 GAAKGLQSLPTHPSPLOQRYSEDPVPLPSETDGV-----APL-----TC-----SPQ 1142
Db      1071 GREQFSRPLSH-PIPRGR-----PASFSSEHVTGSAELQEKVSVCRSRKSISPR 1122
Qy      1143 PE-----YANQPDVPRPQPSPREG-----LPABPACATLERANTLSP-GRNGV 1186
Db      1123 PRGDSAYHSQRHSLTPVTPLSPGLEEDNGYVMDTHLRGASSRSEGLTSSVGLSSV 1182
Qy      1187 V-----KDVFAFGAVENPEVLTPOGGAPOHPHP 1216
Db      1183 LGTEEDDED-----EVEYEMRKRGKSP-PRPP 1209

```

```

RESULT 11
EGFR_DROME          STANDARD;          PRT;      1426 AA.
ID_P04412; 061601; Q9W2G0; P81868;
AC      13-AUG-1987 (Rel. 05, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
       (Gurken receptor) (Tropo protein) (Drosophila relative of ERBB).

```

```

GN      EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS      Drosophila melanogaster (Fruit fly).
OC      Euxaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC      Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxId=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX      MEDLINE=94350209; PubMed=8070664;
RA      Clifford R., Schubach T.;
RT      "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT      that several genetically defined classes of alleles cluster in
RT      subdomains of the receptor protein."
RL      Genetics 137:531-550(1994).
RN      [2]
RP      REVISIONS.
RA      Clifford R., Schubach T.;
RL      Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX      MEDLINE=85124611; PubMed=2982499;
RA      Lynen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT      "The Drosophila EGF receptor gene homolog: conservation of both
RT      hormone binding and kinase domains."
RL      Cell 40:599-607(1985).
RN      [4]
RP      SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC      STRAIN=Oregon-R; TISSUE=Embryo;
RX      MEDLINE=87002474; PubMed=3093080;
RA      Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT      "Alternative 5' exons and tissue-specific expression of the
RT      Drosophila EGF receptor homolog transcripts."
RL      Cell 46:1091-1101(1986).
RN      [5]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP      ANALYSIS.
RX      MEDLINE=99102120; PubMed=9882502;
RA      Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT      "Several levels of EGF receptor signaling during photoreceptor
RT      specification in wild-type, Ellipse, and null mutant Drosophila."
RL      Dev. Biol. 205:129-144(1999).
RN      [6]
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC      STRAIN=Berkely;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherier S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abail J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,
RA      Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beckova K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA      Butts K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Rados B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy U., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,
RA      Reinert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,

```

RA Shih B.C., Stiden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spradlin A.C., Stapleton M., Strong R., Sun E.,
RA Svoboda K., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RP STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Madaworth S.C., Vincent W.S. II; Bjödau-Wentworth D.;
RT "A *Drosophila* genomic sequence with homology to human epidermal
RL growth factor receptor.";
RN Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92036942; PubMed=1936592;
RA Raz E., Schejter E.D., Shio B.Z.;
RT "Interallelic complementation among DER/Elb alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases";
RN Genetics 129:191-201(1991).
RN [9]
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
RT EGF receptor.";
RL Cell 89:13-16(1997).
CC -I- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA
CC AND VENTRAL NEUROECODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III. ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -I- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS, IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC ---
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license@isb-stb.ch. See <http://www.isb-stb.ch/announce/>.
CC or send an email to license@isb-stb.ch.
DR EMBL; AF052754; AAC08536.1; -;
DR EMBL; AF052753; AAC08535.1; JOINED.
DR EMBL; AF052754; AAC08535.1; -;
DR EMBL; AF052752; AAC08535.1; JOINED.
DR EMBL; K03054; AAA51462.1; -;
DR EMBL; K03417; AAA51460.1; -;

	EMBL; K03416; AAA50965.1; -	POTENTIAL.
DR	EMBL; K03418; AAA51461.1; -	EPIDERMAL GROWTH FACTOR RECEPTOR.
DR	EMBL; AF109077; AAD26134.1; -	EXTRACELLULAR (POTENTIAL).
DR	EMBL; AF109078; AAD26132.1; -	POTENTIAL.
DR	EMBL; AF109082; AAD26132.1; JOINED.	CYTOPLASMIC (POTENTIAL).
DR	EMBL; AF109078; AAD26133.1; -	PROTEIN KINASE.
DR	EMBL; AF109078; AAD26133.1; JOINED.	ATP (BY SIMILARITY).
DR	EMBL; AF109079; AAD26130.1; -	ATP (BY SIMILARITY).
DR	EMBL; AF109081; AAD26130.1; JOINED.	BY SIMILARITY.
DR	EMBL; AF109079; AAD26131.1; -	PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
DR	EMBL; AF109083; AAD26131.1; JOINED.	
DR	EMBL; AF109080; AAD26135.1; -	
DR	EMBL; AE003454; AAFA6732.1; -	
DR	EMBL; X02293; CAA26157.1; -	
DR	EMBL; X78920; CAAS5523.1; -	
DR	EMBL; X78918; CAAS5521.1; -	
DR	EMBL; X78919; CAAS5522.1; -	
DR	PIR; A00640; GOFPE.	
DR	HSSP; P11362; TRGX.	
DR	FlyBase; FBgn0003731; Egfr.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; EGF_Kinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_kinase.	
DR	pfam; PF00069; kinase; 1.	
DR	pfam; PF00757; Furin-like; 1.	
DR	pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_kinase; 1.	
DR	SMART; SMO0261; FU_7.	
DR	SMART; SMO0219; TyKc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.	
KM	PROSITE; PSS0011; PROTEIN KINASE DOM; 1.	
KM	Transmembrane Glycoprotein: Receptor; Phosphorylation; Transferase;	
KM	Tyrosine-protein Kinase; ATP-binding; Signal; Alternative splicing;	
KM	Developmental protein.	
FT	SIGNAL 1 30	POTENTIAL.
FT	CHAIN 31 1426	EPIDERMAL GROWTH FACTOR RECEPTOR.
FT	DOMAIN 31 868	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 869 889	POTENTIAL.
FT	DOMAIN 890 1426	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 938 1198	PROTEIN KINASE.
FT	NP_BIND 944 952	ATP (BY SIMILARITY).
FT	BINDING 971 971	ATP (BY SIMILARITY).
FT	ACT SITE 1063 1063	BY SIMILARITY.
FT	MOD_RES 902 902	PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
Query Match	28.1%; Score 1920; DB 1; Length 1426;	
Best Local Similarity	32.5%; Pred. No. 2.7e-95;	
Matches 466; Conservative 184; Mismatches 430; Indels 356; Gaps 44;		
OY	24 QVCTGTDMKRLRPSPETHDMLHLYGCGVVGNGNELTYLP-NMSLSGLODIQEYOG 82	
DB	100 KIGTGSRISLVPSNKHNNLDRTNYCYVDGNIKLTWLPRENIDLSTFLDNI REVTVG 159	
OY	VLIAAHNOVRQVPRRIIVRGTOF-----EDNYVALAVLDNGPRLNNTPTVTGASPGSL 137	
DB	160 YIILSHVDYKKVVPKQIIINGRIITLSLVEEEKTALFV-----TYSIKM 203	
OY	138 RELQLRSITELIKGVLIQRNPOLCYODTLIMKDIFPKNQDLALTLDITNSRACHPCSP 197	
DB	204 YTLIEIPDLRDVNLQGQVFHNHNNYNLCNHRITIOWSIVNGTDAYNNYDTAREERECPKHE 263	
OY	198 MKCKSRCKWESSELCOSTIRVCAAGCA--RCRKLPLTDCCHBOCAAAGCTGPKIASDCLAC 255	
DB	264 SCTHG-CWGGRPKXOCQFKSLTCSPOCAGGCVCPKPRECCHLRCAGGCTGTPROKDIAC 322	
OY	256 LHFHSIGICELEHCALVTYNTDPRESMPNEGRTFPASCYTAPRNUTYSTDVGSCTLYVC 315	
DB	323 KNPFDEAVSVKECPMKRKINTFTYLETNEPGKAAYGATCYKECP-CHLLRDNGACVRSYC 381	
OY	316 PLNHQEVTAIDGTORCEKSKPCARVCYSGLMQYIKANSFFIGITEL-----EFACGCK 369	


```

Db      59  NGSKTPSTIAAGVGGLLCLVAVGIGIYLRRR-HIVKRTLRLLQERLEVEPTPSGE 117
Qy      712  MPNOQMRLKETELRKVKVLSGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANK 771
Db      118  APNOAHRLKETETKVKVLSGSAFGTVYKGLWIPBEKVKI PVAIKELREATSPKANK 177
Qy      772  EILDEAVYMAVGSPYVSRLLIGICTSTVOLVTOQMIPYGCCLLDHVRNKGRLGSDLLNM 831
Db      178  EILDEAVYMAVDNPHVCRLLIGICTSTVOLVTOQMIPYGCCLLDYIREHNDNIGSOYLLNM 237
Qy      832  CMOIAKGSYLEDVLRVLDLAARVNLVKS PNHVITTPGRLRLDIDETEVHAGKVP 891
Db      238  CVQIAKGNVYLERLVRDLAARVNLVKTPOHVITTPGRLAKLGADEKEVHAGKVP 297
Qy      892  IKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQP 951
Db      298  IKMALESILRRITVHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy      952  ICTIDVYMIWKCMWIDSECPRELVSEPSRMARDQRFVVIQ-NEIDLGPASPLDSTF 1010
Db      358  ICTIDVYMIWKCMWIDSDSRKRELIAEFSKMDRPPRYLVIGDERMHLPSPTDSKF 417
Qy      1011  YRSILEDMDGDLVDAEEYLVPOQGFPCPDPAAGAGVNHHRHSSSTRSGGDLTLGLE 1070
Db      418  YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1071  PSEBAPRSP-----APSEAGSDVFPDGLGMAKGLQSLPTHPSPLOKYSDDPVP 1125
Db      450  -----SRTPPLISLSATSNSNATCID-----RNGQGFPRVEDSFVQYSSDPICN 495
Qy      1126  LPSET--DGVAAPLTCSPQEPQVYVNOQDVRPSPREBEPPLAARAGATLEAKTLSGK 1183
Db      496  FLEESIDDFL-----PAPEYVNO--LMPKPS-----TANVQ 526
Qy      1184  NGVYKDV-----AFGAVENPEYVLTPOGGAAPQHPAPSPAPDNLVY 1228
Db      527  NQVNNISLTAISKLPMSRYONSHSTAVDNEYL-----NTNQSPLAKTVESSPY 578
Qy      1229  WDO-----DPE-----RCAPSPSTFKCTPRAENPEYGLDVP 1260
Db      579  WIGSQNHQINLDNPDYQDFLENETKPNGLKVPAAENPEYLRVAP 625

RESULT 13
ERBB_AVIER STANDARD: PRT: 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
   gene family";
   Cell 55:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debutte B., Henry C., Benaisna M., Biserte G., Claverie J.-M.,
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
   new type of oncogene";
   Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC      tyrosine phosphate.
CC      -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC      ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC      IN CHICKENS.
CC      -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC      RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, K02006; AAA42394.1; ALT_INIT.
DR      EMBL, K01216; AAA42400.1; -.
DR      PIR, A00644; TVYU.
DR      HSSP, P11362; 1FCX.
DR      InterPro, IPR000719; Euk_pkinase.
DR      InterPro, IPR001245; Tyr_pkinase.
DR      Pfam, PF00069; pkinase; 1.
DR      ProDom, PD0000001; Euk_pkinase; 1.
DR      SMART, SM00219; TyrKc_1.
DR      PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE, PS50011; PROTEIN_KINASE_DOM; 1.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132..399
FT      NP_BIND 138..146 ATP (BY SIMILARITY).
FT      BINDING 165..165 ATP (BY SIMILARITY).
FT      ACT_SITE 257..257 BY SIMILARITY.
FT      CONFLICT 29..29 S -> W (IN REF. 2).
FT      CONFLICT 140..140 S -> F (IN REF. 2).
FT      CONFLICT 146..146 I -> V (IN REF. 2).
SQ      SEQUENCE 604 AA; 67633 MW; 76EBDD067450609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4,1e-84;
Matches 360; Conservative 76; Mismatches 126; Indels 126; Gaps 16;

Qy      593  CAHYKDPFCVCARCSGVKPPDLSTWPIKFPDEBACQPCPINCITSCVLDLDDKCPAEO 652
Db      3  CAHFIDGPHCVAKCAGVYGENDTL-VRYKADANAVCOLCHPNCITRGCCKPGLGECP--- 58
Qy      653  RASPLTISVAVV-GILLVVVGVVFGILIKRROOKIRKYMRRLLQETELVEPTPSGA 711
Db      59  NGSKTPSTIAAGVGGLLCLVAVGIGIYLRRR-HIVKRTLRLLQERLEVEPTPSGE 117
Qy      712  MPNOQMRLKETELRKVKVLSGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANK 771
Db      118  APNOAHRLKETETKVKVLSGSAFGTVYKGLWIPBEKVKI PVAIKELREATSPKANK 177
Qy      772  EILDEAVYMAVGSPYVSRLLIGICTSTVOLVTOQMIPYGCCLLDHVRNKGRLGSDLLNM 831
Db      178  EILDEAVYMAVDNPHVCRLLIGICTSTVOLVTOQMIPYGCCLLDYIREHNDNIGSOYLLNM 237
Qy      832  CMOIAKGSYLEDVLRVLDLAARVNLVKS PNHVITTPGRLRLDIDETEVHAGKVP 891
Db      238  CVQIAKGNVYLERLVRDLAARVNLVKTPOHVITTPGRLAKLGADEKEVHAGKVP 297
Qy      892  IKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQP 951
Db      298  IKMALESILRRITVHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy      952  ICTIDVYMIWKCMWIDSECPRELVSEPSRMARDQRFVVIQ-NEIDLGPASPLDSTF 1010
Db      358  ICTIDVYMIWKCMWIDSDSRKRELIAEFSKMDRPPRYLVIGDERMHLPSPTDSKF 417
Qy      1011  YRSILEDMDGDLVDAEEYLVPOQGFPCPDPAAGAGVNHHRHSSSTRSGGDLTLGLE 1070

```


Db 418 YRTIMEEDMEDIDVADDEYLPHOGFF-----NSPST----- 449

Qy 1071 PSEEBAPRSP-----APSEGASDVDFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVP 1125
450 -----SRTPLLSLSATSNNNSATNCID-----RAGQGHPRVREDSFVGRYSDDPCGN 495

Db 1126 LPST--DGYVAPLTCSQPEYVNOPDVPRPSPRECEPLPAAPAGATLERAKTSLSPCK 1183
496 FLESISIDGFL-----PAPRYVNO--LMPKKPSTAM----- 524

Qy 1184 NGVVKDVFAP-----GGAVENPEYLTTPCGGAAPQHPPEAPSPAFD 1224
525 --VNOQIYNFISLTAISKLPMDSRKQNSHSTAVDPEYL-----NTNOSPILAKTVFE 574

Db 1225 NLVYWDODPPERGAPPTFKGTPTAENPEY 1254
575 SSPYMIOSGNHQ-----INLDNPDY 594

RESULT 14
ERBB AVIEU STANDARD; PRT; 540 AA.
ID ERBB AVIEU
AC P11273:
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.,
"A single amino acid substitution in v-erbB confers a thermolabile
phenotype to ts167 avian erythroblastosis virus-transformed erythroid
cells.";
RT Mol. Cell. Biol. 6:1751-1759 (1986).
RL -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).

DR EMBL/ M13179; AAA42401.1; -
DR PIR/ A25231; TVEVEB.
DR HSSP/ P13362; IEGK.
DR InterPro/ IPR000719; Euk_Pkinase.
DR InterPro/ IPR001245; Tyr_Pkinase.
DR Pfam/ PF00069; Pkinase; 1.
DR ProDom/ PD000001; Euk_Pkinase; 1.
DR SMART/ SM00219; TYRKC; 1.
DR PROSITE/ PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE/ PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE/ PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN_KINASE.
FT NE BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 ATP (BY SIMILARITY).
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5853297AA068B65D CRC64;

Query Match 23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 2.9e-80;

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

Qy 593 CAHYKDPFCVACRCPGKPLSYWPIWKFPDEBGACQPCPINTCHSCVYDLDDKCGPAG 652
3 CAHFDGHCYKACGAGVLTGENDTL-VKRYADANAACVCLCHNCTRCGCGPGLGCGP 58

Db 653 RASPLTISVAVV-GILLVVLGVVFGILIKRROQKIRKRYTMRLLQETELVEPLTPSGA 711
59 NGSKTPSIAGVGGGLCLLVVVGIGLYLRR-HVKKRTLRLLORELEVEPLTPSGE 117

Qy 712 MPNOQAMILKETELARKKVLGSGAFGVYKGIYIPDEBNKIPLAIKYLRENTSPKANK 771
118 APNOAHILKETEKVKVGFAGFVYKGLWIPBEKVTIPLAIKELRENTSPKANK 177

Qy 772 EILDEAYMAGVSGPVSRILGICLTSTVOVLTOLMPYGCCLLDHRENGRGLSDLLNW 831
178 EILDEAYMASVNDPHVCRLLGICLTSTVOLTLQMPYGCCLLDYREHNDNGSGLLWW 237

Qy 832 CMOIAGKNSYLEDVRLVHRDLAARNVLKSPHNVKITDFGLARLLDIDETEHADGKVP 891
238 CVQIAKGMNVLERRMVRDLAARNVLKTPQHVKITDFGLAKOLGADKEKTHAEGKVP 297

Db 892 IKMALESILRRRTHQSDVWSYGVTVLMTFGAKPYDGIIPAREIPDLLEKGERLPOP 951
298 IKMALESILHRIYTHQSDVWSYGVTVLMTFGSKPYDGIIPASEISSVLEKGERLPOP 357

Qy 952 ICTIDVYIMVKNWIDSECPRELVSESRMARDQRFVVO-NEDLGASPLDSTF 1010
358 ICTIDVYIMVKNWIDSECPRELVSESRMARDQRFVVO-NEDLGASPLDSTF 417

Db 1011 YRSLEDDMDGLVDAEYLYPOGFFCPDPAFGAGVHHRRHRSSTFRSGGDLTLGLE 1070
418 YRTIMEEDMEDIDVADDEYLPHOGFF-----NSPST----- 449

Qy 1071 PSEEBAPRSP-----APSEGASDVDFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVP 1125
450 -----SRTPLLSLSATSNNNSATNCIDRNG-----H----- 476

Db 1126 LPSTDGYVAPLTCSQPEYVNOPDVPRPSPRECEPLPAAPAGATLERAKTSLSPCK 1184
477 -PVREDDGL-----PAPRYVNO--LMPKKPSTAMVNOQIYNFISLTAISKLPIDSRVON 527

Qy 1185 GVVKDVFAPFGAVENPEYL 1203
528 -----SHSTAVDPEYL 539

RESULT 15
EGFR CHICK STANDARD; PRT; 703 AA.
ID EGFR CHICK
AC P13387:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (CER
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.,
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha";
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M20386; AAA48760.1; .
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Puriin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 4.
DR PROSITE: PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE: PS00109; PROTEIN KINASE_TYR; PARTIAL.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703
FT DOMAIN 31 654
FT TRANSMEM 655 667
FT DOMAIN 197 206
FT DISULFID 201 214
FT DISULFID 222 230
FT DISULFID 226 238
FT DISULFID 239 247
FT DISULFID 243 255
FT DISULFID 258 267
FT DISULFID 271 298
FT DISULFID 302 314
FT DISULFID 318 333
FT DISULFID 336 340
FT DISULFID 513 522
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 581
FT DISULFID 589 592
FT DISULFID 592 601
FT DISULFID 605 627
FT DISULFID 630 638
FT DISULFID 634 646
FT CARBOHYD 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON TER 703
SQ SEQUENCE 703 AA; 77427 MM; AFE2DE11B735A690 CRC64;

Query Match 23.0%; Score 1571; DB 1; Length 703;
Best Local Similarity 43.8%; Pred. No. 5.Se-77;

Matches 313: Conservative 113: Mismatches 252: Indels 36: Gaps 15:
QY 8 RWGLLALLPPGAA-----STVCTGTDMLRLPASEPTHLDMLRHLYOGGVQGNLE 61
DB 13 RGAALVYLLGLVALCSAVERKVCQGTNNKLTOLGHNVEDHTSLQRYNNCEVILSNLE 72
QY 62 LTYLPTNASLFLDIOGVGVLIJAHNOVPLORIRYRGTOLEFEDNVALAVDNGD 121
DB 73 ITVEBNRDLTFELKTIOGVAGVLIJALNMVDFLENLQITRGVNLVINDSPALAVSNYH 132
QY 122 PLNNTPVTGASPGRLRELQURSLTEILKGVLIJORNPOLCYODTILMKDIFHKNQAL 181
DB 133 -MNTQ-----GLRELPMKRSLSEILNGGVKISNNPKLQMDVTLVNNDIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKCKSRCKGSESSDQSLTRVYCGGCA-RCKGPLPTCCHEQ 239
DB 183 TLVDFAVSLSSCPKCHPMTCTDHGCGAGEQNCQTLTKVICAQCCSGRGKVPKPSDCHNQ 242
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCPALVNTDTPESMPREGRYTFGASCYTAC 299
DB 243 CAAGCTGRESDBCLACKRFDDATCTKCPPLVLTPTTYQMDVNBKRYSGATCVNEC 302
QY 300 PYNVLTSDVGSCTLVCPILHNOEVTAEDETQCEKSKPCARVYCYGLAQYIKANSKFIGI 359
DB 303 PHNVVTDHGS CVRSCNTDTEV-ENGVKCKCKDGLCSKVCNGIGIGELKGIUS-INA 360
QY 360 TELE-FACCKKIFPSGLAFPSFPGDPASNTAPLOPEOLOVFTLEITGYLYISAWPDS 418
DB 361 TNISFKNCTKINDVSLPVAFLGDAFTKTLPLDPKLDVFRIVKEISGFLLIQAWPDN 420
QY 419 LPDLSVONLOVIRGRIHNGAVSLTLOGLGISWGLRSLRELSGLLIHNNHLCVYH 478
DB 421 ATDLAENLELIRGTRKQGOYSILAVNLKIOSLGLSLKEISDGIJALMKNKLCYAD 480
QY 479 TVPMDOLFRFNNFTVSWFLRVPKVSASHLECEVGEGLACHOLCARGHCMPGPTCVACS 538
DB 481 TMNRSFLPATQS-----QKTKIIONRNKNDCTADRHVCTDPLCSGVCGMGPFHCFSCR 534
QY 539 QFLGQECVEECRVLOGIPREYVNAHRLCPHPECQPNG--SVTCGPEADQVCACAH 595
DB 535 PFSRQKSCVKKCNILQEPREFERSKCLPCHSECLVONSTAYVNTCSGPGDHCMCAH 594
QY 596 YKDPFCVAPRCPGVKPDLSVMPIMKPPDEGACOPCINCTHSCVDLDDGCPAEGQAS 655
DB 595 FIDPFCVAKCPAGVLGBNDTL-VKXTADANAVCOLCHPNCCTGCKGGLGCGP--NGS 650
QY 656 PLTSIVSAV-GLLVVVLGVVFGILIKRQOKIRKTYMRRLQETELVEPLTP 708
DB 651 KTSIAGGVGGLCLVVGIGLGLYLR-R-HIVRKRLRLRLQERLVEPLTP 703

Search completed: July 22, 2003, 08:46:40
Job time : 20.2793 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842
Sequence: 1 MELAALCRWGLLALLPFGA.....TFKGTPTAENPEVIGLDPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_prodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6125	89.5	1259	6	O18735
2	3113	45.5	1209	11	O9QX70
3	3084	45.1	1210	11	O9EP98
4	2719	39.7	1165	13	O9YH40
5	2662.5	38.9	1137	13	O9W6F6
6	2280	33.3	1328	13	P79754
7	2009.5	29.4	1433	5	O9BIF9
8	1871	27.3	419	4	O9UK79
9	1739	25.4	367	11	O8R2X1
10	1720	25.1	729	15	O86712
11	1718	25.1	567	15	O86714
12	1697.5	24.8	412	4	O8WYV0
13	1653.5	24.2	962	15	O64895
14	1645	24.0	545	15	O85468
15	1475.5	21.6	655	11	O9WVFS
16	1459.5	21.3	643	11	O9ERV6

17	1275	18.6	1193	5	O9Y1X8
18	1180.5	17.3	1368	5	O23821
19	1167	17.1	1717	5	O25566
20	1108	16.2	527	13	O9O836
21	1002	14.6	478	11	O9ESE0
22	924.5	13.5	599	13	O9PSH2
23	906	13.2	165	4	O14256
24	806.5	11.8	346	13	P11776
25	797	11.6	176	11	O923V5
26	778	11.4	435	5	O8SZW1
27	754.5	11.0	311	13	O9Y162
28	734.5	10.7	1362	13	O9PV24
29	734	10.6	331	4	O9BUD7
30	723	10.6	149	6	O9BG66
31	721	10.5	1671	5	O9KJVS
32	694	10.1	1418	13	O93457
33	690.5	10.1	1368	13	O8UW85
34	669.5	9.8	1369	13	O8UW86
35	667	9.7	1412	13	O8UW84
36	666.5	9.7	1472	5	O9U5A8
37	654	9.6	1358	13	O73798
38	644.5	9.4	1418	13	O8UW83
39	632	9.2	1245	13	O9YGH8
40	631.5	9.2	2144	5	O9VD94
41	625	9.1	1371	11	O9QVW4
42	598	8.7	987	11	O91YMO
43	593	8.7	987	11	O99MR2
44	591	8.6	935	4	O96L35
45	587.5	8.6	1036	4	O07912

ALIGNMENTS

RESULT 1
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ExB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_taxid=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of exB-2 from canine mammary gland";
RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.
DR HMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; 1PGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00016; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATB-binding; Transferase; Tyrosine-protein kinase.
KW SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;
SQ

O9Y1X8 ephydatia f
O23821 caenorhabdi
O25566 schistosoma
O9O836 gallus gall
O9ESE0 rattus norv
O9PSH2 gallus gall
O14256 homo sapien
P11776 xiphophorus
O923V5 rattus norv
O8SZW1 drosophila
O9Y162 xiphophorus
O9PV24 xenopus lae
O9BUD7 homo sapien
O9BG66 oryctolagus
O9Y1V5 biophalari
O93457 scophthalmu
O8UW85 paralicthy
O8UW86 paralicthy
O8UW84 paralicthy
O9U5A8 bombyx mori
O73798 xenopus lae
O8UW83 paralicthy
O9YGH8 scophthalmu
O9VD94 drosophila
O9QVW4 rattus sp.
O91YMO mus musculu
O99MR2 mus musculu
O96L35 homo sapien
O07912 homo sapien

Query Match	89.5%;	Score 6125;	DB 6;	Length 1259;
Best Local Similarity	89.6%;	Pred. No. 0;		
Matches 1134;	Conservative 44;	Mismatches 76;	Indels 12;	Gaps 3;
Qy	1	MELALALCRKGLILLALLPFGAAS	TOVCTGTDKMLRLPASPETHLMDLRLYGGCVOVGNL	60
Db	1	MELAAWAKWGLILLALLPFGAAGT	OVCTGTDKMLRLPASPETHLMDLRLYGGCVOVGNL	60
Qy	61	ELFTYLPNASTLSFQDIOEV	GVYLIAHNOVROYLORLRYRGTOLEFEDYALAVDNG	120
Db	61	ELFTLPNASTLSFQDIOEV	GVYLIAHNOVROYLORLRYRGTOLEFEDYALAVDNG	120
Qy	121	DPLNNTPTVGTASFGRLQRL	QSLRSLTEILKGVLIQRNPOLCYDPTILMKDIFKNNOLA	180
Db	121	DPLEGGIPARQAAQGLRELO	RLRSLTEILKGVLIQRNPOLCHQDPTILMKDIFKNNOLA	180
Qy	181	LTLIDTRSRACHSCSFMCKGS	RCWGBSSSDCOSLTRVVCAGGACRCGRLPTCCHEQC	240
Db	181	LTLIDTRSRACHSCSFMCKGS	RCWGBSSSDCOSLTRVVCAGGACRCGRLPTCCHEQC	240
Qy	241	AAGCTGPKHSDCLACLFHNS	HCIALVTYNTDTFESMPNDEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACLFHNS	HCIALVTYNTDTFESMPNDEGRYTFGASCVTACP	300
Qy	301	YNYLSTDVSGCTIVCP	LHNOEVTADGTORCEKCSKPCARVCYGLMQYIKANSKFIGIT	360
Db	301	YNYLSTDVSGCTIVCP	LHNOEVTADGTORCEKCSKPCARVCYGLMQYIKANSKFIGIT	360
Qy	361	ELERAGCKKIFGSLAFIPES	FDGDPASNTAPLOPEQLOVFEETLEITGYLYISAMPDLP	420
Db	361	IOEPAGCKKIFGSLAFIPES	FDGDPASNTAPLOPEQLOVFEETLEITGYLYISAMPDLP	420
Qy	421	DLASFQMLQVIRGRIILHNG	AYSLTLOGLIGISMLGRSLRELGSGALIHNTHLCFYNTV	480
Db	421	NLSVFQMLQVIRGRIILHNG	AYSLTLOGLIGISMLGRSLRELGSGALIHNTHLCFYNTV	480
Qy	481	PMQDLFFNNFTVSFWLR	VPKVSASHLEECVGEGLACHQLCARSHCWGPGPTQCVNCSQF	540
Db	481	PMQDLFFNNFTVSFWLR	VPKVSASHLEECVGEGLACHQLCARSHCWGPGPTQCVNCSQF	540
Qy	541	LRGQCEVEGRVLOGLPRE	VYNARHCLPCHRECOPONGSVYCFPREADQCVACAHYDPR	600
Db	541	LRGQCEVEGRVLOGLPRE	VYNARHCLPCHRECOPONGSVYCFPREADQCVACAHYDPR	600
Qy	594	FCVARCPSGVKPDLSFMP	IKWKFADBEQTCQPCINCHSCADLDEKGPAPQASPTISI	653
Db	594	FCVARCPSGVKPDLSFMP	IKWKFADBEQTCQPCINCHSCADLDEKGPAPQASPTISI	653
Qy	661	VSAAVGLILLVVLGV	VFGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRI	720
Db	661	VSAAVGLILLVVLGV	VFGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRI	720
Qy	721	LKETELRKVVYLAGS	AGFTYKGIWIPDGENVKI PVAIKVLRNTSPANKELLIDEAYVM	780
Db	721	LKETELRKVVYLAGS	AGFTYKGIWIPDGENVKI PVAIKVLRNTSPANKELLIDEAYVM	780
Qy	781	AGVGSFYVSRLLGICLT	STVOLVTLMPYGCILDHVENRRLGSGODLLNMCMQIAKMS	840
Db	781	AGVGSFYVSRLLGICLT	STVOLVTLMPYGCILDHVENRRLGSGODLLNMCMQIAKMS	840
Qy	841	YLEDVRLVHRDLAARN	LVKSPNNVKITDFGLARLDOIDEFYHADGKVPDKMALESI	900
Db	841	YLEDVRLVHRDLAARN	LVKSPNNVKITDFGLARLDOIDEFYHADGKVPDKMALESI	900
Qy	901	LRRRFTQSDVWSV	GVTVMEMLTGAKPYDGI PAREIPDLLEKEERLPPOPICITIDYMI	960
Db	901	LRRRFTQSDVWSV	GVTVMEMLTGAKPYDGI PAREIPDLLEKEERLPPOPICITIDYMI	960
Qy	961	MYKCMIMIDSECR	PRFREILVSEFSMABDPORFVVIIONEDLPASPLDSTFYRSLLEDDM	1020
Db	961	MYKCMIMIDSECR	PRFREILVSEFSMABDPORFVVIIONEDLPASPLDSTFYRSLLEDDM	1020
Qy	1021	GDVDAEYLVLPQCG	FFCPDPAPAGAGMWHHRHSSSTRSGGDLTLGLSEBEARPRSP	1080
Db	1021	GDVDAEYLVLPQCG	FFCPDPAPAGAGMWHHRHSSSTRSGGDLTLGLSEBEARPRSP	1080

Db	1014	GDVDAEYLVLPQCG	FFCPDPAPAGAGTAHRRHSSSTRNGGELTLGLSEBEARPRSP	1073
Qy	1081	LAPSEAGSDVFPQD	LDLGMGAAGLQSLPTHDPSLQRYSEDPYPLPSETGGYVALPLTCS	1140
Db	1074	LAPSEAGSDVFPQD	LDLGMGAAGLQSLPSODPSLQRYSEDPYPLPSETGGYVALPLTCS	1133
Qy	1141	POPEYVNPQDV	PPQPPSPREGLPLPAPAPAGATLER-----AKTLPCKNGVVKVDFAFGQ	1195
Db	1134	POPEYVNPQDV	PPQPPSPREGLPLPAPAPAGATLER-----AKTLPCKNGVVKVDFAFGQ	1193
Qy	1196	AVENPEYLTPOGGA	APQHPPEPAPFSPAFDNLVYWDQDPPERGADPPSTKGTPTAENPEYL	1255
Db	1194	AVENPEYLTPOGGA	APQHPPEPAPFSPAFDNLVYWDQDPPERGADPPSTKGTPTAENPEYL	1253
Qy	1256	GLDVPV	1261	
Db	1254	GLDVPV	1259	

RESULT 2

ID	Q9GX70	PRELIMINARY;	PRT;	1209 AA.
AC	Q9GX70;			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE	Epidermal growth factor receptor.			
GN	EGFR.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=LIVER;			
RX	MEDLINE=9025888; PubMed=234246;			
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,			
RA	Earp H.S.;			
RT	"A truncated, secreted form of the epidermal growth factor receptor is			
RT	encoded by an alternatively spliced transcript in normal rat tissue."			
RL	Mo. Cell. Biol. 10:2973-2982(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=LIVER;			
RA	Petch L.A.;			
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=LIVER;			
RA	Guttridge K., Dawson T.L., Earp H.S.;			
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M37394; AAF14008.1; -.			
DR	HSSP; P1362; IFGK.			
DR	InterPro; IPR000494; EGFR_L domain.			
DR	InterPro; IPR000719; Euk_DKinase.			
DR	InterPro; IPR002174; Furin-like.			
DR	InterPro; IPR001245; Tyr_DKinase.			
DR	pfam; PF00757; Furin-like; 1.			
DR	pfam; PF00069; pkinase; 1.			
DR	pfam; PF01030; Recep_L domain; 2.			
DR	PRINTS; PR00109; TYRKINASE.			
DR	ProDom; PD000001; Euk_DKinase; 1.			
DR	SMART; SM00261; Fu_3.			
DR	SMART; SM00219; TyKc; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
DR	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.			
SEQUENCE	1209 AA; 134891 MW; 96FE87F6CC187773 CRC64;			

Query Match 45.5%; Score 3113; DB 11; Length 1209;
 Best Local Similarity 49.9%; Pred. No. 6; je-227;

Matches 641; Conservative 167; Mismatches 353; Indels 124; Gaps 30;

```

Qy 3 LAALCRWGLLALLPPGA--ASTOVCTGDMKRLRLPASBETHLDMRHHYQCGVQGMLE 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 LAALCAAG-----GALBEKVCQGSINRLTQGTDEHDLISQRFNNCEVYLGMLE 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 LTYLPTNASLFLDIOEVQGYVLIAHQVROVPLQRLRYRGTOLEFEDNALAVLNGD 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ITVQRYNDLSFLKTIQEVAGYVLIALTVERIIPLENIQIIRGNALYENTVALAVLSN-- 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 122 PLNNTTPTGASPGGLRELOLRSLTEILKGGVLIQRNPQLCVQDTILMKDI FHKNNOLAL 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 -----YGNKTKGLRELPMKNLOEILIGAVRSNNPLLCMETIIOQRMDIV--QVVFISN 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 182 TLIDTNRS--RACHPCSPMKSGSRCSGSESEDCSLTRTVCAAGCA--RCKGPLPTDCHEQ 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 MSMDVQRHLTGCPKCDPSCPGNSCWGRGEENQKLTKIIICAQGCRKRGSRPSGCCNQ 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 240 CAAGCTGPRHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBSRYTFGASCVTAC 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 CAAGCTGPRSDCLVCHRFDEATCKDTCPLMLYPTTYQMDVNEGYSFGATCVKCC 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 300 PYNLTSTGVSCTLVCPHNOEVTAEQTORCEKSKCARCYGLGMOYIKANKFTGI 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 PRNVVVDHSGCVRACGPDYEV--BEDSVSKCKCDGCFKRCVNGIIGEFK--DTLSTINA 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 360 TELE--FAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLOVEETLEETLYLYIACMPD 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 TNIHFKKYCTAISGLHLPLVAFKGDSTFTRPRDLPRELEIKYKEITGFLIIAMWEN 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 419 LPDLSVFONLOVIRGRIIHNGASYSITLQGLGISWGLRSRLBLSGLALIHNNTHLCFVH 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 WTDLHAFENLEIIRGRTKQHGQFSLAVVGLNITSIGLSRKEISGDVYISGNRRLCYAN 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 479 TVPMDOLFRPNFPTYSFWLRYPKVSASHLE--ECVEGGLACIQGLCARHGCMGPRQCVC 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 TINMKKLGFTN-----QKTKIMNNAEKDCATNHCPLCSBECGWGPEPTDVCSC 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 538 SQPLRGCEVECECRVLOGLPREYVNAHCLPCHPECPONGSVTFCGEADQCVACAYKK 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 QNVRGRRCVDKCNILBEGPREFVENSEICHPCEPLQTMNITGTGGRPNICAHYV 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 598 DRPCCVARCEGVCVPRDLSYMPYIKFPDEEGACQPCPNCTHSCVDLDKXCPAEGRASP- 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 DGPCHVKTCPSGINGENNTL--VMKFADANNVCHLCHANCTYGCAGPGLKGC--QQPEGPX 643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 657 LTSIVSAVAGILVAVGVGVFI--LIKROOKIRKYMRRLLOETELVEPLTPSGAMNQ 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 IPSIATGIVGGLFIV--VALGIGLPMRRDQVRKTRRLLOERELEVEPLTPSGEAPNQ 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 716 AQMRILKETELRKVYLGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTSPPKANKETILD 775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 AHLRLKETEFKKIKVYLGSAFGTVYKGIWIPDGENVKI PVAIKVLRENTSPPKANKETILD 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 776 EAYMAGVGSPPYVRLGICLTSTVQVLTQIMPYGCLLDHVENRNGRLGSDLLMNCQI 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 EAYVMASVDNPHVCRLLGICLTSTVQVLTQIMPYGCLLDYVREHKNIGSQYLLMNCQI 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 836 AKGWSYLEDVRLVHRDLAARVVLKSPHVKITDQGLARLLDIDETETHAAGKVPYKMM 895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 AKGNMYLEDRLVHRDLAARVVLKSPHVKITDQGLARLLDIDETETHAAGKVPYKMM 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 896 ALBSILRRPTQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGEPLPOPICTI 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 ALBSILRRPTQSDVMSYGVTVWELMTFGSKPYDGI PASISISILKGEPLPOPICTI 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 956 DVTYIMKCMWIDSECRPRFRELVSERSMARDPQRFVVIQ--NEDLGPASLSTFYKSL 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 DVTYIMKCMWIDSECRPRFRELVSERSMARDPQRFVVIQ--NEDLGPASLSTFYKSL 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1015 LEDDMDGDLVDAAEEYLVQOGFPCRPDPAKAGMWHHRHSSSTRSGGDLTLGLEPSEE 1074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 MEEDMEDVDVDADEYLVQOGF-----NSPST----- 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 1075 BAPRSLPABSGAGSDVFDGDLGMGAAGLQSLPTHDPSPILOYSEDPTVPSPSER--DG 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1031 --SRTPPLSLISANSN-----SSTVACINRNSCRVKEDEAFLORYSDSPFVLTEDNIDRT 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1133 YVAPLTSPQPEYVNPQVPRPQPPSPREGPLPAAPRAGATLERAKTSLSPKNGVVKDVA 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 FL-----PVPEYINQ--SVPRKPAQSVQNPVYHNPPLHP-----APGRDLHYQN--P 1127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1193 FGAVNPEYVL--TPCGGAAPQHPHPPAFSPAFNDNLYYNDQ-----DP-----PE 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1128 HSNAVSNPEYVNTAQ-----PTCLSSGFSLSALMIQKSHQMSLDNPDYQODFFPK 1178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1236 RGAPPSTFKGTPTAENPEYLGIDVP 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1179 EAKPNGIFKG--PTAENAYLKVAPP 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 3

```

Q9EP98
AC Q9EP98 PRELIMINARY; PRT: 1210 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ AND 129/SVEVYAC.
RA Reiter J.L., Threagill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sincalir C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threagill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sechnl C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAC28045.1; -.
DR EMBL; AF275364; AAC28045.1; JOINED.
DR EMBL; AF275365; AAC28045.1; JOINED.
DR EMBL; AF275367; AAC24386.1; -.
DR HSSP; P11362; IFK.
DR MGD; MGI:95294; Egfr.
DR InterPro: IPR000345; CytC heme bind.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Btk_kinase.
DR InterPro: IPR002174; Pyrin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PFO1030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Btk_kinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Tyrc; 1.

```

DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 KM ATP-binding; Receptor; Transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 45.1%; Score 1084; DB 11; Length 1210;
 Best Local Similarity 49.4%; Pred. No. 1e-224;
 Matches 632; Conservative 169; Mismatches 359; Indels 120; Gaps 27;

QY 11 LLLALLPFGAA--STVCTGTDMKRLRLPASPETHLMDLRHLYOGCOVVOGNLELYLPTN 68
 DB 14 LLTLLCAAGALAEKKVCOGTSNRLTQGTGFEDHFLSLQRYMNCVVLGNLEIYYGRN 73
 QY 69 ASLSFLDIOEVQGVLLAHNQVQVPLQRLIYRGTOGLFEDNVALVLDGDLNATTP 128
 DB 74 YDLSFLKTIQEVAGVLLIALNTVERIPLLENLQIIRGNALYENTVALAILSN----- 124
 QY 129 VTGASPGRLRELQRLSTEILKGVLIQRPOLCYODTILMKDI----FKNNQLATLTI 184
 DB 125 -YGNRRGLRLPRLNLOEILIGAVRFNNPILCMMDITIQMRDLYQVNFNMSMDL--- 180
 QY 185 DTNRSRACHPCSPMCKSGRCSGESSEDCSLTRTVACGCA-RCKGPLPTDCHEQCAAG 243
 DB 181 -QSHPSGCPKCDPSCPNCSWGCGEENCOKLTIIICAQCSHRGGRGSPDCCHQCAAG 239
 QY 244 CTGPKHSDCLACHFNHSGICELHCPALVTYNTDFFESMPREGRYTGACVTAAPVNY 303
 DB 240 CTGPRSDCLVCOKEODEATKDTCPPLMLYNTPTTYQMDVNPBEGYSGATCVKCKPNY 299
 QY 304 LSTVGSCTTIVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIRANSKFIGITELE 363
 DB 300 VVTHGSCVVRACGPDYEV-EDGIRKCKKCDGPRKCNIGIGIEFR-DLISNATNIK 357
 QY 364 -FACCKKIFGSLAPESFGDDPASNTAPLOEQLQVEFLEITGYLYISAMPDSELDL 422
 DB 358 HFKCTAISGDLHLPVAFKGDSEFTRPPLPRELEIKTYKEITGFLIIQAMPDMDTL 417
 QY 423 SVFONLOVIRGRILHNGAVSLTLQGLISWGLSLRLSGLGLLIHNNHLCFVHTYPM 482
 DB 418 HAFENLEIRKTYQHGFSLAVGGLNTSLGLSLKEISQGDVITISGRNLCAVANTINW 477
 QY 483 DOLFRFNFTVSMILRVKVSASHLE-ECVGEGLACHQLCARGHGCGPPTQCNVCSQFL 541
 DB 478 KKLFGTNN-----QKTKIMNNRAEKOCKAVNNHYCNPLCSGECGWEPRDCTSCQVVS 530
 QY 542 RGDCVEECRYLQGLPREYVNAHCLPCHPECPQNGSVTCFGRPADCVACAHYKDPF 601
 DB 531 RGRECVKCNILBEPREFVNSSECIQCHPECLFOAMNITCTGRGPDNCICAHYIDPH 590
 QY 602 CVARCPGSGVPRDLGYMPIWKFDEGACQPCPINCTHSQVLDLCKGCAEGRASPLSTIV 661
 DB 591 CVKTCFPGMIGENNTL-VMKYADANNVCHLCHACTTGACGPGLOGCEVWPSGKPIBIA 649
 QY 662 SAVVGILLVVLGVVFGI-LIKRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNQAQMRI 720
 DB 650 TGIIVGLLFTIV-VALGIGLFMRBRHLYRKXTRLRLLOERLVEPLTPSGAPQAHLRI 708
 QY 721 LKETELARKVYLGSGAGFTYKGIWIPDEGENVKIPVAIKVLRNTSPKANEILIDEAYVM 780
 DB 709 LKETEFKRIKVLGSGAGFTYKGLMIPDEGEVKIPVAIKELREATSPKANEILIDEAYVM 768
 QY 781 AGVGSFVYSRLIGLCTSTVOLVTLQMPYGLLDHVENRGRLSODLLMNCQIAGMS 840
 DB 769 ASVDNPHVCHLLGICLSTVOLITQMPYGLLDHVENRGRLSODLLMNCQIAGMS 828
 QY 841 YLEEDVRLVARDLAARNVLYVSPNHVKITDFGLARLDIDETEVHADGKVPYIKMALESI 900
 DB 829 YLEDRRLVHRDLAARNVLYVTPQHVKITDFGLAKLLGAEEKVEYAEAGKVPYIKMALESI 888
 QY 901 LRRRTHQSDVMSGVTVWELMTGARGPYDGIPIAREIPDLLEKEERLPORPCTIYVMI 960
 DB 901 LRRRTHQSDVMSGVTVWELMTGARGPYDGIPIAREIPDLLEKEERLPORPCTIYVMI 960

DB 889 LHRITYHQSDVMSGVTVWELMTGSKRPYDGIPIASDISILEKEERLPORPCTIYVMI 948
 QY 961 MYKCMIDSECRPPREFRELVSFSKMARDPQRFVYIQ-NEDLGAPSLDSTFYSRLLEDD 1019
 DB 949 MYKCMIDSECRPPREFRELVSFSKMARDPQRFVYIQ-NEDLGAPSLDSTFYSRLLEDD 1008
 QY 1020 MGDVLDAEYLVPOGFFCPDPAPGAGGMVHRRSSSTRSGGGLTGLSESEEAR 1079
 DB 1009 MEDVVDADDEYLTLPQGGFF-----NSPST-----SKT 1034
 QY 1080 PLASBEGAGSDVDFDGLMGAAKGLQSLPTHDPSPLGRSEDPVPLPSET--DGYVAPL 1137
 DB 1035 PLLSLSATSN-----NSVACINRNGSCRVVEDAFLOKRYSSDPGAVTEDNIDDAFL--- 1087
 QY 1138 TCSPPQPEYVNOPDVPPRPPREGLPAPARFAGATLEAKTSLPGKNGVAVKDVAFGGA 1197
 DB 1088 ---FVPEPVYNO-SVPKBPAGSVQNPVHYNOPLHP-----APGRDLHYON--PHSNV 1133
 QY 1198 ENPEYL-TPQGGAAPQPHPPAPFAPAPDNLYVMDQ-----DP-----PERGAP 1240
 DB 1134 GNPEYLVNTAQ-----PTCLSSGFVSPALMTQKSHQMSLDNPDYQODFPFKETYPN 1184
 QY 1241 STFGKTPFAENPEYVGLDVP 1260
 DB 1185 GTFKPG-PTAENAEYLRVAP 1203
 RESULT 4
 Q9YH40 PRELIMINARY; PRT; 1165 AA.
 ID Q9YH40;
 AC Q9YH40;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphia; Acanthopterygii; Percomorphia; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duesch J.,
 RA Altschimed J., Scharcl M.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.",
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Scharcl M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U534771; AAD10500.2; -.
 DR HSSP; P11362; IRGK.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Dkinase.
 DR InterPro; IPR002174; Puriin-like.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR Prodom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYKc; 1.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE; PS00143; GRAM_pos_ANCHORING; UNKNOWN_1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 2.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; kinase; Transferase; Tyrosine-protein kinase.
 SO SEQUENCE 1165 AA; 129614 MW; 7f7EE38D871A74E CRC64;

Query Match 39.7%; Score 2719; DB 13; Length 1165;

Best Local Similarity 45.4%; Pred. No. 4,5e-197;
 Matches 582; Conservative 163; Mismatches 383; Indels 154; Gaps 31;

```

QY 1 MELALCMGILLALPPG-AAST-----QVCTGTMKRLPASPETHLDMRLHYGQCV 55
DB 4 LELLEL-----LILLLLISGCCSTDPDRKVCQGTSNQMTM--LDNHLKMKKMSGCNV 56
QY 56 VQGNELTYLPTNASTFLQDIOEVGYVLIANNOVPLQRLIRVGTOLFEDNYALA 115
DB 57 VLENLEITYTOENQDLSLOSTQIEVGYVLIANNEVSTIPLVNLRLINGONLYEKNFL 116
QY 116 VLDNGDPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIQNPOLCYQDTILMKDIFRK 175
DB 117 VMSNVQK-NPSSP--DYVQVGLKQLSLNLTFLISGVYVSHNPLCNVERITNMWDIVDK 173
QY 176 NNQALATLIDNRSACHPCSPMCKSGSRCKGSSSDQSLRTVCAGGC-AACKPLPTD 234
DB 174 TSNPTMLNIPRAFERQCCQCDPGCVNCSQWAPGCHCQFTYLLCAEQCNRCRCRKPID 233
QY 235 CCHQCCAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPGRYTFGAS 294
DB 234 CCHQCCAGCTGPRATDCLACDFPDQDTCQTCPPKITYDVSHQVVDNPIKTFGAA 293
QY 295 CVTACPVNYLSTDVGCTVLPLHNOEYTABDGTORCEKSPKRCVCGYGLM---QYI 350
DB 294 CVKCPSPNYVTE--GACVRSACAGMLEVD--ENGRKSKCRPCDQCVKCDGIGISLSMTI 351
QY 351 KANSKFTIGITELFPAQCKKIFGSLAFLPESFPGDASTAPLOPQLOVFETLEITGYL 410
DB 352 AVNSTNIG----SFSNCKTNGIDILNRSNFEQDHYKIGPMDPHLMNLTIVKKEITGYL 407
QY 411 YISAMPDSLPLDSYVQNTQVIRGRILHNGAYS-LTLQGLISMLLSRLSELGSLALIH 469
DB 408 VIMMPEMTSLSYQNLLEIRGRITTFSRGFSFVVVQVSHLOMLLSLKEVASNVIK 467
QY 470 HNTHLCFVHTVPWDLFFRNFTVSPMLRVKVSASHLECEVGEGLACHOLCARGHCWP 529
DB 468 NTPQLRVASTINMRLFRSEDSQIEYDART-----ENQTCNNESEDGCGP 514
QY 530 GPTQCVNCSQFLRGECEYBEGRVLOGLPREYVNAHCLPCHECCOPNGSVTCFPEADQ 589
DB 515 GPTMCVSCLAHYDRGRCVASCNLLQGBEPREAVDGRCVQCHQDECLVQTDLSLTCYGPGRAN 574
QY 590 CVACAHYDPPFCVAPRCPSGVYVPLDSTYMPIMKFPDEEGACOPCPINCHSCVVDLDKGP 649
DB 575 CSKCAHFDGQPCIRCPHGMIGDDTL-IMKRYADMKQCCOPCHQNTQCGSGPPLSGCR 633
QY 650 AEORASPLTIVSAVVGILLVVLGVFGLIKRQOKIKRKYMRLLQETELVEPLTPS 709
DB 634 GD-IVSHSSLAVALSGLLITVYVALLVLLRRRIK-RKKTIRLLQEKELVEPLTPS 691
QY 710 GAMPRQADMRILKETELKRVKVLGSGAGFTYKGIWIDGDEVNKPVALIKVIRENTSPKA 769
DB 692 GOAPRQAFRLIKETELFKDRVLGSGAGFTYKGIWIDGDEVNIRIPVALIKVIRENTSPKV 751
QY 770 NKEILDEAYVWAGVSPVYSRLGLCTSTVOQLVQLMPYGLDHFVENRGRGLSQDL 829
DB 752 NQEVDAEAYVWASVDPHVCRLGLCLTSAYQVLQMLPFGCLDHFQDRICGQML 811
QY 830 NMCQIAKMSYLEVDVRLVHRDLAARNVLYVSPNNVKTITDFGLARLDIDETEHADGK 889
DB 812 NMCVQIAKMSYLEVDVRLVHRDLAARNVLYLKNPNVKTITDFGLARLDADEKEYAGHGK 871
QY 890 VPIKMALESTLRRFTHQSDVMSYGVVWELMTGAPRYDGIPIAREIPDLLEKEERLPQ 949
DB 872 VPIKMALESTLQWYTHQSDVMSYGVVWELMTGSKRYDGIPIAREIASVLENGERLPQ 931

```

```

QY 950 PPICTIDVYIMVWKCMIDSECRPRELVSFSSMARDPORFVYIIONEDLCPASPLDST 1009
DB 932 PPICTIEYIMLKKMIDPSSRPFRFLVGEFSGMARDPKRYLYIQ---NLPSPSDR 968
QY 1010 FYSRLIEDDMGDLVDAEYVLPQGFPCPPAPGAGCMHHRHSSSTRSGGDLTLGL 1069
DB 989 LFSRLISSDD--DVVDADAEYLL-----RYKRIN-RQGS----- 1018
QY 1070 EPSEEARSPPLAPSEGSQSVFQDGLGMAKGLQSLPTHDPSPLQRYSEDPV-PLPS 1128
DB 1019 -----EPCIPNGH-----PVRENSTALKRTISDPONALEK 1049
QY 1129 ETDGVAPLTCSPQPEYVNOQDVRQP-----PSPR-GLP-AARPAGATLER 1175
DB 1050 DLDGH-----EYVNOQGSSTSRSLSDIYNVNYEDLTDGKCPVLSLQEAETNFSR 1099
QY 1176 AKTISPGNGVAVQVFAFGAVENDEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPE 1235
DB 1100 PEYLTNQNSSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQ 1134
QY 1236 RGAPSTFKGTPTAENPEYGL 1257
DB 1135 TGAITGNGMFLPAENLEYGL 1156

```

RESULT 5

```

QY 09W6F6 PRELIMINARY; PRT; 1137 AA.
ID 09W6F6
AC 09W6F6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN ERBB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HINDRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "Distribution of neurogulin-1 (ng1) and erb4 transcripts in
  RT embryonic chick hindbrain."
  RL Mol. Cell. Neurosci. 13:237-258(1999).
  DR EMBL; AF121963; AAD31764.1; -.
  DR HSSP; P11362; 1FGK.
  DR InterPro; IPR000494; EGFR_L_domain.
  DR InterPro; IPR000719; Euk_pkinase.
  DR InterPro; IPR002174; Furin-like.
  DR InterPro; IPR001368; TNFR_C6.
  DR InterPro; IPR001245; Tyr_pkinase.
  DR InterPro; IPR004019; VLP_motif.
  DR Pfam; PF00757; Furin-like; 1.
  DR Pfam; PF00069; pkinase; 1.
  DR Pfam; PF01030; Recep_L_domain; 1.
  DR Pfam; PF02757; VLP; 2.
  DR PRINTS; PR00109; TYRKINASE.
  DR Prodom; PD000001; Euk_pkinase; 1.
  DR SMART; SM00261; Fu; 3.
  DR SMART; SM00219; Tyrc; 1.
  DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
  DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
  DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
  KW PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
  KW kinase; Tyrosine-protein kinase.
  FT NON TER 1
  SO SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC94F CRC64;

```

Query Match 38.9%; Score 2662.5; DB 13; Length 1137;

Best Local Similarity 46.3%; Pred. No. 8.4e-193;
Matches 533; Conservative 171; Mismatches 352; Indels 95; Gaps 28;

```

Qy 161 LCYDTIMKDIIFHKNQALTLIDTNRSPACHPSPMCKSRCKGSESDCQSLTRVC 220
Db 3 LCFADTHIMODIVRNPMASNFTLVPTNGSSGGRCHKSCGTG-RCMGPTENHQQTLTKTV 61
Qy 221 AGGC-ARCKGLPTDCCHECCAGCTGPKHSDCLALPHNHSIGICELHCPALVTYNTDTF 279
Db 62 AEGCDGRCTGYGVSDCCRECGAGCGSGPRDIDCFKCMFNDSGACVTCQPGTFVNPPTF 121
Qy 280 ESMENPEGRYTFGASCVTACPNYLSTDVSGCTLVCPHNOEVTADGTORCEKSKPCA 339
Db 122 QLEHNHAKYTYGAFCKKCPHNFF-VDSSECVRACPSKMEV-EENGKMKCKPTDIDCP 179
Qy 340 RVCGGLGMOYIKANSKFIGITELE-FACCKKIFGSLALPESFDDPSNAPLOPEQLO 398
Db 180 KACGIGIGTSL-VSAQTVDSSNIDKFINCTKINGLIFLVGIGHDPYHTIAIINPEKLN 238
Qy 399 VFETLEETGYLYISAMPDSLPLDSVFQNIYIRGRIIHNGAYSLTLQIGISWGLRSL 458
Db 239 IFQIVREITGYLNIOSPEENMDRFVSNLVTIGRALYSGLSLILKQOQITSLQFSL 298
Qy 459 RELSGSLAIHNTLCEVHTVPMQDLRFNNFTVSFWLRVPKVA-----SHLEEYGE 513
Db 299 KQISAGNIYITDNSLCYHTVNTSLF-----STPSQKTVIHRNKKXKENCITAD 347
Qy 514 GLACHOLCARHGCWGPGTQCVNGSQPIRGQECVECVLQGLPREYVNAHCLPCHPC 573
Db 348 GMVNCNELCSSDCMGPGPDQGLSCRFIRGRICIESCNLYGGEFREFANGSCMECDQC 407
Qy 574 QP-ONGSVTCFPGREADOCVACAHYKDPRECVARCSGVKPDLSYMPKPFPEBEGACPC 632
Db 408 EKEMDMNTTCGPGPDHCTCFHFDGPNVCYCKCDGLQGANF--IKRIADEDECHPC 465
Qy 633 PINTHSCVDLDDKCC-----PAQRASPLTISVSAVV-GILLVVVLGVVFGIL 680
Db 466 HPNCTGCGRASHDCIYYPMTROSTLPHAR-FTL--IAAGVIGLPIYIMGLTFAYV 522
Qy 681 IKRQOQKIRKXTMRLLLOETELVERLTPSGAMPNOQMRILKETELRKVYLGSAGFTV 740
Db 523 VRRKSIK-KKRALRFL-ETELVERLTPSGTAPNOAQRIIKETELRKVKVLGSAGFTV 580
Qy 741 YKGIWIPDGENVKIPVAIKVLRENTSPRANKELIDEAVYAVGVSPVSRLLGLTSTV 800
Db 581 YKGIWIPDGENVKIPVAIKILNETTGPANVEPHEALIMASMDPHVLRLGVLSTPTI 640
Qy 801 QLVTLMPYGLLDHVRNKGRLSGQDLLNMCQIAKMSYLEDVRLVHRDLAANVLVK 860
Db 641 QLVTLMPHGLLDVYHKNIDNGSOLLNMCVOIAKGMVLEERLVRDLAANVLVK 700
Qy 861 SPNVKLTDFGLARLLEDEKEYNADGKMPKMMALCIIHYRKRTHSDVMSYVITME 920
Db 701 SPNVKLTDFGLARLLEDEKEYNADGKMPKMMALCIIHYRKRTHSDVMSYVITME 760
Qy 921 LMTGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMIWIKCMIDSECRPRELVS 980
Db 761 LMTGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMIWIKCMIDSECRPRELVS 980
Qy 981 EFSMARDPORFVVIQNEE-LGPASPLDSTFYRSLLEDDMGDLVDAEYLVPOGFPCCP 1039
Db 821 EFSMARDPORFVVIQNEE-LGPASPLDSTFYRSLLEDDMGDLVDAEYLVPOGFPCCP 1039
Qy 1040 DRAGAGGMVHHRSSSTRSGGDLTLGLEPSEEAR--PLAP-SEGAGSDVFDDDL 1096
Db 880 PPIYTSRTRIDSNRNOFYRGGYAAEGV-FMPYRAGCIIPEAPVQAQTAELFEDTC 938
Qy 1097 GMAKAGLQSLPTDPSPLQRYSEDPVPLPS-----ETDGAAPLTCGPOPEYVNP 1149
Db 939 CNGTLRKQVATLAKEDSDSTORYSADPTVFIPERVIRGLDLBDGVTTPRKDKPDYLPV 998
Qy 1150 DVRRQPSRPGRLPA-RPAGATLERAKTSLSPGNGVAVKVF-----AFGGAENP 1200

```

Db 999 EENPFVSRKKNQDLQAVDNPEYH-----APNGCPRAEDEYVNEPLYLNTFANTLEVA 1051

```

Qy 1201 EYLTPQGAAPQPPRPAFNFADNLYTWDQPPRGA--PPSTKGRPT----- 1248
Db 1052 EYL-----KNMPEKAKKAFNDPDYMNHSLPPRSTLOHPDYLOEYSTKYFYKNGRI 1103
Qy 1249 -----AENPEYL 1255
Db 1104 RPIYANPEYIL 1114

```

RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.

```

ID P79754
AC P79754
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ErpB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; Fu; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

```

Query Match 33.3%; Score 2280; DB 13; Length 1328;
Best Local Similarity 39.9%; Pred. No. 1.1e-163;
Matches 516; Conservative 162; Mismatches 417; Indels 198; Gaps 33;

```

Qy 9 WGLLLALLLP--GAASDQ-----VCGTDMKRLRPAPEPTLMDLRLHYOGCOVQGNLEL 62
Db 4 WRLILMCVARSRLRAASSSTQDAVCGTONGISSTGSOENQYLNLDKRYKGGIINGNLEI 63
Qy 63 TYLPTNASTFLQDIOEVQGYVLIAHNOVRPQLRLRIVNGTQLFEDNYALVALVDNGDP 122
Db 64 TQISNMDFSLKTRIREVTGYVLIAHNFQRIPIGLQKLVINGNSLYERFALSVFLN--- 120
Qy 123 LANTTPVTGASPGGLRELQSLRLEILKGVLIQRPOLCYQDITLWKDIFHKNQALALT 182
Db 121 ----YPKG--PSGILNQLGMLNLTLEIDGVOIINNKRLRYGPMWYWRDII-RNNDADIE 173
Qy 183 LIDTNRSPACHPSPMCKSRCKGSESDCQSLRTVAGGC-ARCKRPLPTDCCHECA 241
Db 174 IQNGERGCVH---KSC-GATCMGPGKQCOILTKTVACAPQCNDRCFSTSPDCCHICA 229
Qy 242 AGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFESMNPBEGRYTFGASCVTACPY 301
Db 230 AGCKGRLPTDCAFRLFNDSGACVPGCPQTLIYKQFQWETNPNKAYQSGSICVSGCPT 289

```



```
QY 302 NYLSTDVSCCTLVCPNLHNOETAEADGTOR-CEKSKPCARCYGLGMOYITANSKFIET 360
DB 290 HFV-VDSSSCVSVCPDMEV--ERGSQOCELCSCGCPKVCCEGAG--ORQTVDS 342
QY 361 ELE-FACCKIIFGSLAFLPESFDGDPASNTAPLOEOLQVETLEIITGYLYISAMPSL 419
DB 343 NIDFINCTKIQGSLHFLVTGILGDDFNVPPLDAKLEVRITREITDIANTOSMPREL 402
QY 420 PDLSPFQNLQVIRGRILHNGVSLTLOGLSIWLGRSLRELSGLALIHNTHLCFVHT 479
DB 403 NDLVSFSLTTIQRSILFKRPSLWMRIPLTLSLGRSLREISQSVYISQNALHCYHT 462
QY 480 VPMQDLRFNNFTVSFHLRVPKVSASH-LEECVGBGLACHOLCARGHWGCGPPQCVNCS 538
DB 463 VNMVQLFRGSR-----VRANSLSNRMACEVADGRVCDLPSPSGCMGCGPPQCLSCR 516
QY 539 OFLHGCEVCECRVLQGLPREYVNAH-CLPCHPECPQONSVTYCFGEADQCVACANHYK 597
DB 517 NYSHGTCVAGCHNSGIPREFAGLVGCVACHPECKQTOKASTGFGADECMACTKFR 576
QY 598 DPPECVACPSGVKPDLSYMPIMKFPDEGACOPCPINTHSQVLDKCGCPARQASPL 657
DB 577 DGPYCMSSCPAGVN-DGEKGLIFKFPNREGHCEPCHQCTGCGSGPLNDC---LEAARL 632
QY 658 TSIVSAVVGILLVYVGVF-----GILKROQKRTKRTMRLLQETLVEPLTSG 710
DB 633 TISSGQTTGIALGVPAGLIFCLVLPFLGMLVHRLAIRRKAMRYLESSEGSFBLGP-G 691
QY 711 AMPQAOAMRILKETELRKVKVYLSGAFGTYYKGIMIPDEGVNKPVAIKVLRNMTSPAN 770
DB 692 EKGKVARILKPSDLKIKPLGSGVGTGSKGFWIPGEVTKIPVALKITIDSSGRQTF 751
QY 771 KEILDEAYVAVAGVSPYVSLGLICLTSTVOLVTQLMPEYGLLDHVENRGRLSQDLN 830
DB 752 TEITDHLTSMQSLDHPYIVRLGLICPGTCLQTLVQLSHSGSLLEHIRQKTSLDPQLN 811
QY 831 WCMQIAGMSYLEVRLVHRLAARNLVKSPNHYKITDPELALLDIDETETIADGKV 890
DB 812 WCVQIAGMYLLESHRVHKNLAARNLLKNQYQVQISDGVADLLVPDDKYYVSEKTK 871
QY 891 PIKMALESILRRRFTHQSVDVSYGVYTWELMTFGARVYDGI PAREI PDLLEKGERLPOP 950
DB 872 PIKMALESILFRRTYHOSDVSYGVYTWEMMSFGAEYASVQOEVPVLEKGERLSQP 931
QY 951 PICTIDVYIMVYKCMWIDSECRPRFREIVSEFSRMAPDQRFVYIIONEDLGASPLDSTF 1010
DB 932 AICTIDVYVWVYKCMWIDENIRPTFEKELASDFTWADPRLVYLRMEG----- 980
QY 1011 YRSLLEDDMDGLVDAEYLYVPOGFCPPAPAGGMVHHRRSSSTRSGGDLTLGLE 1070
DB 981 -----EDSGMGEFL-----RRGSR-----GILLEADLE 1003
QY 1071 PSEBEARSPPLAPSEAGSDVFDGDLGMG---AAKGLQSLPTHPSPLQ-----R 1117
DB 1004 EDEBE-----GIGDRFAITPSLOPSPSSTWSQINSYVMTQLR 1042
QY 1118 YSEDPVLPBSETDGYVAPLTCSPQ- EYVNO-----PDVAPQPSPREP 1162
DB 1043 YD-----FAVSGGHI GYLPMSPSPYDITRLQVYQSRSLSSVYTLDRSAFRSSSEAE 1096
QY 1163 L--PAARPAATLERATKLSPGKNVYKDVAFAGCAVENPEYLTPOGGAAPQHPAPAS 1220
DB 1097 LCEBDGACAGIFRVR-----FQSEKGN-----PQCG----- 1122
QY 1221 PAFDNLVYWDOPPERGAPSTFKGTPTAENPE 1253
DB 1123 -----QORKLSTASSPSSFTMADEDE 1146
```

```
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxId=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SDA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ01655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Puriin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON TER
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.4%; Score 209.5; DB 5; Length 1433;
Best Local Similarity 32.4%; Pred.No. 3.9e-143;
Matches 473; Conservative 194; Mismatches 366; Indels 409; Gaps 39;

QY 26 CTGIDMLRLRPASETHLDRLHLYOGGVYQGLTYLPTNMSLSFLQDIQVQGYVL 85
DB 1 CIGTNGMSVYANREHYKMLRDRTYNTCTYVDGMLITWQNTIDNLFQHIRVETGYVL 60
QY 86 IAHNVQVPLQRLRIVRGTLF-----EDNVALAVLDNGDPLNNTPTVTGASFGGLREL 140
DB 61 ISLYDLPOVILPRQITIRGRTTFKLNMKEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGVLIQNPOLCYODITLWMDI-FHKNQQLATLITDNRSPACHPCSPMC 199
DB 105 ELPLRLDILGGSVGFANNYMLCHMKSLNMBELILAPQTSMTYNTFSPSPREAVCPCHSC 164
QY 200 KGSFCKESSEBDQSLRTTYACAGCA--RCKGRLPTDCHEQCAAGCTGPRHSCLACLH 257
DB 165 EVG-CWGEAGNCCRFKSLNCSPOCSQGRGPRRCCHLFCAGGCTGTPQSDCLACKN 223
QY 258 FNSHGICELCPALVYNTDTFESMPNPEGYTGASCVTAACPYNVYSTDVGSCITVCP 317
DB 224 FYDDGVKQCEFPQOINPTNYFEPNPDGRTAYGATVAKCP-EHLIKDNGACVRKCPK 282
QY 318 HNOEVTAEADGTORCEKSKPCARCYGLGMOYITANSKFIETLEFAGCKIIFGSLAFL 377
DB 283 GKMPQNSE-----CVPKGVCPKTCPEGI-----VHSDNIG-----NYKDCITIIIGSELEIL 329
QY 378 PESFDGPPASNT-----APLOPEOLOFETLEETITGYLYISAMPSLSPDLSPVONLO 429
DB 330 DOSFDFQOQVYTFNSFGPRYIKIDPRLLEVFSVKEITGFINIOAHHPNFTLLNVPNNLE 389
```



```

RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match
Best Local Similarity 88.0%; Pred. No. 1.7e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRTTHGSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRTTHGSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60
QY 955 IDVMIMVWCKMMISSECRPRRELVSERSMARADQRFVIVIONEDLGFSFPMDSFFYSLSL 1014
DB 61 IDVMIMVWCKMMISSECRPRRELVSERSMARADQRFVIVIONEDLGFSFPMDSFFYSLSL 120
QY 1015 LEDDDMGDLVDAEEXLVLPQGFPCPDPAAGAGVNHHRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDAEEXLVLPQGFPCPDPAAGAGVNHHRSSSTRSGGDLTLGLEPSEE 180
QY 1075 EAPRSPLAPSGAGSDVFDGDLGMAAGLQSLPTHDSPLOKRSSEDTVPLPSTDGIV 1134
DB 181 EAPRSPLAPSGAGSDVFDGDLGMAAGLQSLPTHDSPLOKRSSEDTVPLPSTDGIV 240
QY 1135 APLTCSPOPEYVNOQDVVPOPSPREGLPARAPAGATLERAKTSLPKNGVVKQVAFG 1194
DB 241 APLTCSPOPEYVNOQDVVPOPSPREGLPARAPAGATLERAKTSLPKNGVVKQVAFG 300
QY 1195 GAVENPEYLTPOGGAAPQHPHPAPFSPAFDNLVYWDQDPERGAPESTFKGPTAENDEY 1254
DB 301 GAVENPEYLTPOGGAAPQHPHPAPFSPAFDNLVYWDQDPERGAPESTFKGPTAENDEY 360
QY 1255 LGLDVVPV 1261
DB 361 LGLDVVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
ID 086712
AC 086712;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Polypeptin.
GN POLYPEPTIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL: S69372; AAC60725.1; -.
DR HSSP: P03322; 1A6S.
DR InterPro: IPR000719; Euk Dkinase.
DR InterPro: IPR004026; Retro M.
DR InterPro: IPR001245; TYR_kinase.
DR Pfam: PF00069; Dkinase; 1.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

```

```

DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914FE1D63 CRC64;

Query Match
Best Local Similarity 54.8%; Pred. No. 1.3e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PONGSVTCFGEADOCVACAHKYDPPFCVAPCGSGVKPLSYMPIMKFPDEGACQPCPI 634
DB 141 PEETPTPTKTP--DHCMCAHFIDPCHVCYACPAVLGENDTL-VKVIADANAACVCLHP 197
QY 635 NCTHSCVDLDKXGCPAEQASPLTISVAVV-GILLVVLGVVFGILLKROQKIRKTYM 693
DB 198 NCTRCCKPGLEGCP---NGSKTSPISIAAGVVGILLCLVVGIGIGILYLR--HIYKRTL 253
QY 694 RRLDQETLVEPLTSPSGAMPNOAKRILKETELRKYVLGSGAFGTVYKGINIPGSENK 753
DB 254 RRLDQETLVEPLTSPSGAMPNOAKRILKETELRKYVLGSGAFGTVYKGINIPGSENK 313
QY 754 IPVAIKVLRNTSPKANKELIDEAYVMAGVSPYVRLGICLTSTVOLVTLMPYGCIL 813
DB 314 IPVAIKELRENTSPKANKELIDEAYVMASVDNPRVCRLLGICLTSTVOLVTLMPYGCIL 373
QY 814 DHVENRGRGLSODLLNMCOTAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 873
DB 374 DYIREHKONISQOYLNNVCVOAKGMNTEBRRLVHRDLAARNVLKTPQHKITDFGLA 433
QY 874 RLIDIDEFTHADGKVPILKMALESILRRFTHOSDWSYGVTVWELMTFGAKPYDGI 933
DB 434 KLLGADKEXYHAEGKVPIKMALESILRRITTHOSDWSYGVTVWELMTFGSKPYDGI 493
QY 934 AREIPDLLEKBERLPQPICTIDVMIMVWCKMIDSECRPRRELVSERSMARADQRFV 993
DB 494 ASEISSVLEKBERLPQPICTIDVMIMVWCKMIDADRPKRELIDERSKMARADQRFV 553
QY 994 VIQ-NEDIGPASPLDSTFYRSLLEDDDMGDLVDAEEXLVLPQGFPCPDPAAGAGVNHHR 1052
DB 554 VIQDERKHLPSPTDSKYFRTLMEEDMEDIDVDAEEXLVLPQGF----- 598
QY 1053 HRSSSTRSGGDLTLGLEPSEBEARSPPL-----APSGAGSDVFDGDLGMAAGLQSL 1107
DB 599 -NSPST-----SRTLPLSLSATSNNSATNCID-----RNGQH 631
QY 1108 PTHDPSPIQRYSEDTVPLPSET--DGVAAPLTCSPOPEYVNOQDVVPOPSPREGLPAR 1165
DB 632 PVREDSFYQRYSSDPTGNFLBESIDDFL-----PAPEYVNO--LMPKPS----- 675
QY 1166 ARPAGATLERAKTSLPKNGVVKQV-----AFGAVENPEYL 1203
DB 676 -----TAMVONQIYNNISLTAISKLPMSDRYQNSHSTAVIDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
ID 086714
AC 086714;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";

```

RL Oncogene 9.1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; -.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase.1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyrcK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR Tyrosine-protein kinase.
 DR NON TER
 FT 1
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
 Best Local Similarity 55.4%; Pred. No. 12e-121;
 Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQVCAHAYKDPFCVACPSGVKPDLSYMPIMWFPEDEGACOPCPINCTHSVDL 643
 DB 1 GP--DHCKKCAHFDGPHCVKACPAVIGENDTL-VWKYADANAVCQLCHPCTGCKGP 57
 QY 644 DDKGCPAEQASPLTSIVSAVV-GILLVVLGVFGLIKRQOKIRKRYMARLLQETEL 702
 DB 58 GLEGCP--NSKTPSIAGVGVGLCLVVGGLGILYLR--HIVRKRTLRLLQEREL 113
 QY 703 VEPLTPSGAMPNOQMRILKETELRKVYLGSGAGTYKGIWIDGENVKIPVAIKVL 762
 DB 114 VEPLTPSGEAPQAHRLILKETEFKRVKVLGSGAGTYKGIWIDGENVKIPVAIKELR 173
 QY 823 ENTSPKANKEILDEAVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGLLDHVENGR 822
 DB 174 EATSPKANKEILDEAVYVAGVSDNPRVCRLLGICLTSTVOLVQLMPYGLLDYIREHND 233
 QY 823 LGSQDLMLWCMQIAKMSYLEDVRLVHRDLAARNVLVKSPPHVKITTFGLARLLIDETE 882
 DB 234 IGSQDLMLWCMQIAKMSYLEDVRLVHRDLAARNVLVTPHVKITTFGLAKLLGADKE 293
 QY 883 YHAGQKVPKIMMALESILRRFTHSDVMSYGVTVWELMTFGAPYDGIPIARELPDLE 942
 DB 294 YHAEQKVPKIMMALESILRRFTHSDVMSYGVTVWELMTFGSPYDGIPIARELSVLE 353
 QY 943 KGERLPOPICTIDYVMIMVCMIDSECRPRELVSFESRMARDPQRFVVIQ-NEDIG 1001
 DB 354 KGERLPOPICTIDYVMIMVCMIDADSRPRELVSFESRMARDPQRFVVIQ-NEDIG 413
 QY 1002 PASPLDSTFTYRSLLEDMDGDLVDAEYLVPOQGFCCDPAPAGAGMVAHRRSSSTRSG 1061
 DB 414 LPSEPTDSKFYRSLMEEDMEDIVDAEYLVPHOGF-----NSPST--- 454
 QY 1062 CGDLTLGLEPSEEEAPRPL-----APSEGAGSDVFPDDLGMAKGLQSLPTHDPSPLO 1116
 DB 455 -----SRTPLLSSLSATSNNSATNCTID-----RNGQGHVVRDSFVQ 491
 QY 1117 RYSEDPTVPLPSET--DGTVAPLTCSPQPEYVNOQDVAPQPPSPREGPLPARPAGATLE 1174
 DB 492 RYSDPTCNFLPESIDGFL-----PAPEYVNO--LMPKKPS----- 526
 QY 1175 RAKTLSPGKGVVDVF-----ARGAVENTPEYL 1203
 DB 527 ----TAMVQNIYNNISLTALSKLPMDSRYONSHSTAVDNPEYL 566

RESULT 12

Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (Tremblrel. 20; Created)
 DT 01-MAR-2002 (Tremblrel. 20; Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21; Last annotation update)
 DE Hypothetical 44.7 kDa protein.

GN P31659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OK NCBI_Taxid=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF318349; AAL55856.1; -.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyrcK; 1.
 DR PROSITE; PS00108; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 DR Hypothetical protein.
 KW SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 2.8e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRRFTHSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLKEGERLPOPICT 954
 DB 1 MALESILRRRFTHSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLKEGERLPOPICT 60
 QY 955 IDVTMIMVCMIDSECRPRELVSFESRMARDPQRFVVIQ-NEDIGPASPLDSTFTYRSL 1014
 DB 61 IDVTMIMVCMIDSECRPRELVSFESRMARDPQRFVVIQ-NEDIGPASPLDSTFTYRSL 120
 QY 1015 LEDDDMGDLVDAEYLVPOQGFCCDPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEE 1074
 DB 121 LEDDDMGDLVDAEYLVPOQGFCCDPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGY 1134
 DB 181 EAPRSPAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGY 240
 QY 1135 APLTCSPQPEYVNOQDVAPQPPSPREGPLPARPAGATLEERAKTLSPGKGVVDVFAFG 1194
 DB 241 APLTCSPQPEYVNOQDVAPQPPSPREGPLPARPAGATLEERAKTLSPGKGVVDVFAFG 300
 QY 1195 GAVENPEYLTPOGGAAPQ-----HPPA--FSFAFDNL 1226
 DB 301 GAVENPEYLTPOGGAALSPULLPSAOPSTSTIGTRTHSGGLHPAPSKGLRQSTOST 360
 QY 1227 YVMD-ODPPER-----GAPSTFKGTPTAEN 1251
 DB 361 VVWTCQCEPBGQVARSPPVSSGREGULSAGIKXWEGPPTTSRGTCHARN 410

RESULT 13

Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (Tremblrel. 01; Created)
 DT 01-NOV-1996 (Tremblrel. 01; Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 01; Last annotation update)
 DE Gag, v-erb-A, v-erb-B protein.
 GN Gag, v-erb-A, v-erb-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OK NCBI_Taxid=11861;
 RN
 RP SEQUENCE FROM N.A.

RX MEDLINE=90206603; PubMed=1969616;
 RA Brinkin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transactivating potential of the oncogene v-erb-B.";
 RL Oncogene 5:15-24(1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X52209; CA36459.1; -
 DR EMBL; X52211; CA36459.1; JOINED.
 DR HSSP; P10828; ZNBL.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000536; Hormone_rec_1ig.
 DR InterPro; IPR001723; Stechmann_receptor.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR001628; Znf_Casteroid.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00105; zf-C4; 1.
 DR Pfam; PRO0398; STRDHORMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR ProDom; PD000035; Znf_Casteroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00219; TykC; 1.
 DR SMART; SM00399; Znf_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; DNA-binding; Nuclear_kinase; Receptor;
 KW Transcription regulation; Transferase; Tyrosine-protein kinase;
 KW Zinc-finger.
 QO SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B95CE CRC64;

Query Match	24.2%	Score 1653.5	DB 15	Length 962
Best Local Similarity	51.7%	Pred. No. 2.1e-116		
Matches 358	Conservative 73	Mismatches 142	Indels 119	Gaps 18
Qy	547	VEECRVLOGLPRE-YVNAR-HCLP-----	CHPCQC	574
Db	354	IEKQESYLLAFEHYINVRKGINHHFWSKLMKLVADLRMIGAYHARSRLPHMKVECPETLS		413
Qy	575	PONGSVTCFGEADQCAAHYKDPFCVARNCPSGVKDLSYMPIWKFPEDEGACOPCI		634
Db	414	POE-----VGP--DHCMCAHFHIGPHCVKACAPAGVLENDTL-VMKYADANAACQLCHP		465
Qy	635	NCHTSYVDLDKCGPAEGRASPLTSYSAVY-GILVVVLGVGSGILIKRQOKIRKTYM		693
Db	466	NCTRGCGKPGLEGCP--NGSKPTSIAGVVGGLCTLVVGLIGLTYLR--HIVRKTL		521
Qy	694	RRLQETELVEPLTPSGAMPQAOMRIKETELRKVKVLSGASGAGTVYKGIWPDGEVYK		753
Db	522	RRLQDEBELVEPLTPSGEAPVQAHRIUKETEFKKVKVLSGAGSTYKGLMTPGEKVT		581
Qy	754	IPVAIKVLENTSPKANKEILDEAYVYVAGVSPYVSRLLGICLTSTVOLVQLMPYGCIL		813
Db	582	IPVAIKELREATSPKANKEILDEAYVYVAVSDVNPVYCRLLGICLTSTVOLVQLMPYGCIL		641
Qy	814	DYHRENGRGLSGODLLWMCQOIAKMGSTLEDBVLVHRDLAARNVLYVSPNHVKTIDFGIA		873
Db	642	DYIREHNDNGISQYLLNWCVOIAKMGNYLEERHVAHRDLAARNVLYVTPQVKTIDFGIA		701
Qy	874	RLLDIDETEHADCGKVPKIMMALESLRRFTHOSDQWVSGVYVWELMTGCAPIYDIP		933
Db	702	KQLGADKEVTHAEGKVPKIMMALESLIHRITYTHOSDQWVSGVYVWELMTGSKRPGIIP		761
Qy	934	AREIPLDLKEGERLPORPITCIDVYIMIMVKCMITDSECRPRFRELVSFEMADDPORV		993
Db	762	ASEISSVLEKGERLPORPITCIDVYIMIMVKCMGASDRPRFRELIAFSMAADPPRYL		821
Qy	994	VIO-NEDLPASPLDSTFYFNSLEDDMDGLVDAAEYLVPOQGFCDPAPAGAGVYHR		1052
Db	822	VIOGDERMHLPSIPDPSKFYRTLMEEDEMDIVDADEYLVLPPOGFF-----		866

```

2y      1053 HRSSTREGGCDLTGLGLEPSEEAERPSFLAPSEAGSDVFPDGLGMGAALQSLPTHTDP 1112
          |||
          867 -NSPST-----SRPPLSLSLATSN-----NSATKCIDNNGH-- 898
Db
Qy      1113 SPLQRYSDFTVPLPSETDGVVAVLTCSPQREYVNOQDVRRQPSPREGLPAARPAQAT 1172
          |||::|
          899 -----PVREDGFL-----PAREYVNO--LMEKKPSLAWNOQIYVYISLT 936
Db
Qy      1173 -LERAKTLSPGKGVVNDVFAFGAVENPEYL 1203
          ::::|
          937 AISKLPMDSRVYQ-----SHSTAVDNPYL 961
Db

```

Query Match	Best Local Similarity	24.0%	Score 1645;	DB 15;	Length 545;
Matches 345;	Conservative 70;	Mismatches 121;	Indels 92;	Gaps 15;	
QY	584	GPADOCVACAHYKDPFCVCARCPGSGVKNPLSYMPIKPFDEEGACPCPCINCTHSCVDL	643		
Db	1	GP--DHCMKCAHFIDGPHCVKACGAGVLGENDTL-VKQADANAACVQLCHPNCCTCGCKGP	57		
QY	644	DDKGCRAEORASPLTISVSAVY-GILLVVVLGVVFGILLKRRQOKIRKTYMRRLLDTEFL	702		
Db	58	GLEBCP---NGSKTPSIAGVGVLGLLTVVVGAGIGLYLERR-HIVKRTLRRLDLOREL	113		
QY	703	VEPLTPSGAMPNOAKMRLKETELRKQVLGSAFGVYVNGIMIPGSENVKIVAIKVLK	762		
Db	114	VEPLTPSGAMPNOAHRLKETELPKKVLGFGAFGVYVGLMIPGSEKVTIPAIKELR	173		
QY	763	ENTSPKANKEILDEAYVMAGVSGSPVYSRLGLICTSTVOLTLQIMPGCLLDHVRNCR	822		
Db	174	EATSPKANKEILDEAYVMASVDNPHVCRLGLICTSTVOLITQLMPYGLLDVIRHKDN	233		
QY	823	LGSGODLLNMCQIAKGMSTYEDVRLVHRDLAARVLYKSPNNHYKITDFGLARLLDTEFE	882		
Db	234	IGSQYLLNMCVQIAKGMNVLDEEHLVHRDLAARVLYKTPQDYKITDFGLAKOLGADKE	293		
QY	883	YHADGCVPIKMAALSLIRRFTHSDWMSYGVATWELMTPGAKPYDGIPAEIPDILE	942		
Db	294	YHADGCVPIKMAALSLIRITTHSDWMSYGVATWELMTPGSKPYDGIPIASELISVSE	353		

```

OY 943 KGERLPDPPICTIDVYIMVWCMWIDSECRPRFRELVSFMSMARDPQRFVIO-NEDLG 1001
DB 354 KGERLPDPPICTIDVYIMVWCMWIDSECRPRFRELVSFMSMARDPQRFVIO-NEDLG 413
OY 1002 PASPLDSTFFYRLLEDDEDDMGDLVDAAEYIVLPQOQFPCPPAPAGAGMHHRRSSSTNSG 1061
DB 414 LPSPTDSKFYRLTMEEDMEDIVDADEYIVPHQGF-----NSST----- 454
OY 1062 GDDLTLGLEPSEEEAPRSP-----APSEAGSDVFDGDLGMAKGLQSLPTHPSPILQ 1116
DB 455 -----SRTPLLSLATSNSATNCLDRNG-----H----- 481
OY 1117 RYSEDPTVPLPSETDGVVAPLTCSQPEYVNOPDVRPOPSPREGPLPAAPAPAGAT-TER 1175
DB 482 -----PYREDEGFL-----PAPEYVNO--LMPKKPSTAMVQIQYVYISLTATSK 523
OY 1176 AKTISPGKNGVKKVFAFGAVENPEYL 1203
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

```

RESULT 15

O9WVFS PRELIMINARY; PRT; 655 AA.

AC O9WVFS; 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).

DE EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

[1]

SEQUENCE FROM N.A.

STRAIN=C57BL/6J; TISSUE=LIVER;

RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

Maibhe N.J.; "Alternative transcripts from the Human and Mouse EGFR Genes Encode

RT Cdkoxy-terminal truncated Receptors.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.

STRAIN=C3H/101, 129/SVJ; AND 129/SVEYTAC.

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Fearall R.S., Green P.J., Yee D., Lampland A.L.,

Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

Maibhe N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

RT mouse alternative Egfr transcripts encoding truncated receptor

isoforms.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE FROM N.A.

STRAIN=C57BL/6J; TISSUE=LIVER;

MDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Toh M., Ishii Y.,

Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamashita I.,

Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,

Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

Kuehl P., Lewis S., Marcu S., Nikaido I., Peeble G., Quackenbush J.,

Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,

Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

Brownstein M.J., Bulc C., Fletcher C., Fujita M., Giardioli M.,

Gusiewicz S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

Lyons P., Marchionni L., Mashima J., Marzarelli U., Mombetters P.,

Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Wetz C., Whiteaker C., Williams L.,
RA Wymshaw-Boris A., Yoshida K., Haegawa Y., Kawai H., Kohsuke S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL, AF124513; AAD4149.1; -.
DR EMBL, AF275366; AAG28047.1; -.
DR EMBL, AF275365; AAG28047.1; JOINED.
DR EMBL, AF275365; AAG28047.1; JOINED.
DR EMBL, AK004944; BAB23688.1; -.
DR EMBL, AK004944; BAB23688.1; -.
DR EMBL, AK004911; BAB23662.1; -.
DR MGD: MG1:95294; Egfr.
DR InterPro; IPR004944; EGFR_L_domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

Query Match 21.6%; Score 1475.5; DB 11; Length 655;
Best Local Similarity 43.7%; Pred. No. 3.8e-103;
Matches 283; Conservative 99; Mismatches 232; Indels 33; Gaps 11;

```

OY 11 LLLALPFGAA--STQVCTGDMKRLPASPTLMDMLHYOGCOVVOGNETLYLPTN 68
DB 14 LLLALCAAGALAEKVKVCGTSNRLTQGTEDHFLSLQRMVNEVVLGNLEITYVORN 73
OY ASLFLDIOEVGVVLAHNOVROPLQRLRYVGTQLFEDNVLAVLNDNDPLNNTTP 128
DB 74 YDLSLTQIYVAGVYLAALNTVERIPLENQIITGNALYENTYALALISN----- 124
OY 129 VTGASPGGLRLQLRLSTELIKGVLIOENPOLCYDITLWKDI----FKKNQLATLI 184
DB 125 -YGTNRTGLRELPRNLQELILGAVFSNNPILCMMDITIQRDIVQVFMMSMDL--- 180
OY 185 DTNSRAPHPCSPMCKSGRCMGESSEDDQSLTRYVACAGCA-RCKGRLPTDCHEQCAAG 243
DB 181 -QSHPSGCPKCDPFCPNKSGWGEENCKLTKIKICAOQCHRCGRSPSCCHNQCAAG 239
OY 244 CTGPHSDCLACLHFNHSGICELHCPALVTYNTDFFESMPDEGRYTGASCVTACPNY 303
DB 240 CTGPRESDCLVQCFODEATKDTCPPLMLYNPITYQMDVNPBGKTSFGATCVKCCPNY 299
OY 304 LSTDVGSCTVCPPLHNOEVTADGTORCEKSKPCARVYGLGMQYIYANSKFTIGTELE 363
DB 300 VVTDHGSQVRAAGCPDYEV--BEDGIRKCKKDDGCRKVCNIGIGEFK-DTLISATNIK 357
OY 364 -FACCKIFGSLALPESFDGDPASNTAPLOPQLOVETLEITGYIYISAMPDSDLPL 422
DB 358 HFXYCTAISGDHLIPVAFKDSFTPTPRLPRELEIKYKEITGFLIIQAMPDNNMTL 417
OY 423 SVFONLOVIRGRIHANGVSLTLQGLISWGLSLRLSRLSGLLIHNHTLCPFNHTVPM 482
DB 418 HAFENLEIRKTRTQHQGFSLAVGLNTISGLSLSEISGDIYISGNRLCYANTINW 477
OY 483 DQLFRFNNFTVSWLRAVPKVSASHLE-BCVGEGLACHQLCARGHCWGPPTCVNCSQFL 541
DB 478 KKLFGTRN-----QTKIKINNRAEDCKAVNHVNCPLGSSBECGWGPEDVCSQVNS 530
OY 542 RGQCVCEKRYLQGLPREYVNAHCLPCHPCQPNQSVTCFGEADQCVAAHYKDPF 601
DB 531 RGRCEVCKNLTLEBPPEFENSECICHPCCLPQANNITCTGGRPDNCLQCAHYIDGPH 590
OY 602 CVARCPGKVPDLSYMPIMKPPDEGACQCPRICTHSCVDLDKGC 648
DB 591 CVKTCPRAGINGENNTL-VKXADANNVCHLCHANCTTGACGPGLOGC 636

```

Search completed: July 22, 2003, 09:01:28
Job time : 54.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds

(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-465-479-14

Perfect score: 6842
Sequence: 1 MELALALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6632	96.9	1255	21	AAV92620 Human heregulin 2
2	6632	96.9	1255	22	AAE12130 Human tyrosine kin
3	6632	96.9	1255	22	AAE60167 HER2 transgene pla
4	6632	96.9	1255	23	AAU74545 Human HER2 (ErbB2)
5	6626	96.8	1255	17	AAW01111 HER-2/neu protein.
6	6626	96.8	1255	20	AAW92406 Human HER-2/neu on
7	6626	96.8	1255	21	AAAB21198 Human HER-2/neu pr
8	6626	96.8	1255	21	AAV84780 Amino acid sequenc
9	6626	96.8	1255	22	AAE84548 Human HER-2/neu pr
10	6626	96.8	1255	22	AAE88267 HER2/neu amino aci

11	6626	96.8	1255	23	AAE24067 Human Her-2 protei
12	6626	96.8	1255	23	AAE20479 Human Her-2/neu pr
13	6626	96.8	1255	23	AAE51143 Human Her-2/neu on
14	6626	96.8	1255	23	AAU77114 Human Her-2/neu po
15	6583	96.2	1433	14	AAE39568 Sequence of c-erbB
16	6462	94.4	1223	23	AAU98923 Human breast cance
17	6309	92.2	1200	21	AAE21208 Human HER-2/neu pr
18	5851.5	85.5	1256	21	AAE21199 Rat HER-2/neu prot
19	5851.5	85.1	1256	21	AAE21199 Rat HER-2/neu onco
20	5825.5	85.1	1256	21	AAE21206 Mouse Her-2/neu pr
21	5825.5	85.1	1256	22	AAE2860 Amino acid sequenc
22	5825.5	85.1	1256	23	AAE51151 Mouse Her-2/neu on
23	4712	68.9	919	21	AAE21203 Human HER-2/neu fu
24	4712	68.9	919	23	AAE51148 Human Her-2/neu ex
25	3975.5	58.1	920	23	AAE51152 Mouse Her-2/neu ex
26	3975.5	58.1	926	23	AAE51153 Mouse Her-2/neu ex
27	3596	52.6	712	21	AAE21204 Human HER-2/neu fu
28	3596	52.6	712	23	AAE51149 Human HER-2/neu fu
29	3450	50.4	782	18	AAE19764 Her2-GW-CSF immuno
30	3448	50.4	653	21	AAE21200 Extracellular HER-
31	3448	50.4	653	23	AAE51145 Human Her-2/neu on
32	3410	49.8	645	22	AAE60408 Human ErbB2 extrac
33	3410	49.8	645	22	AAE60408 Human ErbB2 extrac
34	3345	48.9	951	21	AAE41993 DCSctv-erbB2EC fu
35	3242	47.4	624	21	AAE08222 Human EGF receptor
36	3118	45.6	1210	21	AAE19259 Human acid sequenc
37	3118	45.6	1210	21	AAE50616 Human EGF receptor
38	3118	45.6	1210	23	AAE23019 Human Her-1 protei
39	3118	45.6	1210	23	AAE50768 Human epidermal gr
40	3116	45.1	1210	22	AAE68420 Amino acid sequenc
41	3084	45.1	583	23	AAE20483 Human protein for
42	3084	45.1	587	23	AAE20481 Human protein for
43	3083	45.1	589	23	AAE20484 Human protein for
44	3083	45.1	600	23	AAE20482 Human protein for
45	3077	45.0	1210	23	ABP51768 Human epidermal gr

ALIGNMENTS

RESULT 1					
AAV92620					
ID	AAV92620	standard; Protein; 1255	AA.		
XX	XX				
AC	AAV92620;				
XX	XX				
DT	10-AUG-2000	(first entry)			
XX	XX				
DE	Human heregulin 2 (Her2).				
XX	XX				
KW	Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;				
KW	self-protein; cancer; breast cancer; prostate cancer;				
KW	cell-associated peptide antigen; foreign epitope.				
XX	XX				
OS	Homo sapiens.				
XX	XX				
FX	FX				
FT	Domain				
FT	FT				
FT	FT				
FT	Region				
FT	FT				
FT	Region				
FT	FT				
FT	Region				
FT	FT				
FT	Region				
FT	FT				
FT	Region				
FT	FT				
FT	Domain				

Location/Qualifiers
1..173
/label= N-terminal
/note= "mature polypeptide"
5..25
/label= insertion region
/label= "suitable for foreign epitope insertion"
59..73
/label= insertion region
/note= "suitable for foreign epitope insertion"
103..117
/label= insertion region
/label= "suitable for foreign epitope insertion"
149..163
/label= insertion region
/label= "suitable for foreign epitope insertion"
174..323

FT	Region	/label= Cysteine_rich_domain
FT	210..224	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	250..264	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	324..483	
FT	/label= ligand_binding_domain	
FT	325..339	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	369..383	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	465..479	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	484..623	
FT	/label= Cysteine_rich_domain	
FT	579..593	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	624..654	
FT	/label= Transmembrane_domain	
FT	632..652	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	653..667	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	655..1010	
FT	/label= Tyrosine_kinase_domain	
FT	661..675	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	695..709	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	710..730	
FT	/label= insertion region	
FT	/note= "suitable for foreign epitope insertion"	
FT	1011..1235	
FT	/label= C-terminal_domain	
XX	WO200020027-A2.	
XX	13-APR-2000.	
XX	05-OCT-1999;	99WO-DK00525.
XX	05-OCT-1998;	98DK-0001261.
XX	20-OCT-1998;	98US-0105011.
XX	(MEBI-) M & E BIOTECH AS.	
XX	Steinaa L, Moulitsen S, Nielsen KG, Haaning J, Leach D, Dalum I,	
XX	Gautam A, Birk P, Karlsson G;	
XX	WPI; 2000-349917/30.	
XX	N-PSDB; AAA09455.	
XX	Inducing immune responses to weakly immunogenic, tumor associated	
XX	peptide antigens for the treatment of breast and prostate cancer	
XX	Claim 62; Page 193-198; 220pp; English.	
CC	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues	
CC	Her2 can be used in the claimed method as an autovaccine to induce a	
CC	response. Subdominant CTL epitopes, antibody binding regions and	
CC	cysteine residues involved in disulfide bonds are preserved in the	
CC	immunogenized forms. Regions suitable for the insertion of foreign	
CC	helper epitopes were identified (see features table). The method	

	is used for inducing immune responses against weakly immunogenic
CC	cell-associated peptide antigens (PA) such as those associated with
CC	cancers (self-proteins), e.g. human prostate specific membrane antigen
CC	(PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC	The method comprises effecting simultaneous presentation by antigen
CC	producing cells (Apep) of the animals immune system of: (1) at least 1
CC	CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC	B-cell group derived from the cell-associated PA; and (2) at least 1
CC	filter T helper cell group which is foreign to the animal. Analogues of
CC	human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC	part of all known and predicted CTL and B-cell epitopes of the respective
CC	PA and including at least one foreign T helper epitope are also claimed.
CC	The method is used to treat prostate, prostate/breast or breast cancer
CC	when the PA is human PSM, FGF8b and Her2, respectively.
XX	
SQ	Sequence 1255 AA;
Query Match	96.9%; Score 6632; DB 21; Length 1255;
Best Local Similarity	97.2%; Pred. No. 0;
Matches 126;	Conservative 8; Mismatches 21; Indels 6; Gaps 1
Dy	1 MELALATRWGILLALLPPGAASVTCGTDMKRLPASPETHLMDLRHLYGGCVOVGNNL 60
Db	1 MELALATRWGILLALLPPGAASVTCGTDMKRLPASPETHLMDLRHLYGGCVOVGNNL 60
Dy	ELTLPNPNASLSFLDDIOEVGVYVLIAHNOVRVPDLRLRVRGTOLEFDNYALAVALDNG 120
Db	ELTLPNPNASLSFLDDIOEVGVYVLIAHNOVRVPDLRLRVRGTOLEFDNYALAVALDNG 120
Dy	DPLNNTTPVTGASPDGLRELDRLSLTEILKGCVLIQRNPOLCYODITLMXDIFFHKNNOLA 180
Db	DPLNNTTPVTGASPDGLRELDRLSLTEILKGCVLIQRNPOLCYODITLMXDIFFHKNNOLA 180
Dy	LTLIDTRRSAPCHPCSPMKGSRWGSSESDCOGLTFTVCAGGCARCKGPLPTDCHEOC 240
Db	LTLIDTRRSAPCHPCSPMKGSRWGSSESDCOGLTFTVCAGGCARCKGPLPTDCHEOC 240
Dy	AAGCTGPKHSDDLACLHPNHSGICELCPALVTYNTPFPESMPPEGRYTFGASCATACP 300
Db	AAGCTGPKHSDDLACLHPNHSGICELCPALVTYNTPFPESMPPEGRYTFGASCATACP 300
Dy	YNTLSTDVGSCTLYCPAHNOBVTAEADGTORCEKSKPCANVCYGLGMQIVKANSKFITG 360
Db	YNTLSTDVGSCTLYCPAHNOBVTAEADGTORCEKSKPCANVCYGLGMQIVKANSKFITG 360
Dy	ELEPAGCKTKFGSLAFIPESFDGPANNTAPLOEGLOVETTELIEITGYLTISAMPDSLP 420
Db	ELEPAGCKTKFGSLAFIPESFDGPANNTAPLOEGLOVETTELIEITGYLTISAMPDSLP 420
Dy	DLVSFONLQVIRGRILHNAGAYSLTLOGLGISWILGRSLREIGSGLAIHHNTHLCFHYTV 480
Db	DLVSFONLQVIRGRILHNAGAYSLTLOGLGISWILGRSLREIGSGLAIHHNTHLCFHYTV 480
Dy	PWOLFPRFNPFYSFMLRVPKVASASHLEECVSGSLACHOUCARGHCMPGPGTCCVNCOSF 540
Db	PWOLFPRFNPFYSFMLRVPKVASASHLEECVSGSLACHOUCARGHCMPGPGTCCVNCOSF 540
Dy	LARGOECEBEERVLQGLPRELYNARHLCPHCEPOCPONGSYTCGFEPADOCVAACHYDXDP 600
Db	LARGOECEBEERVLQGLPRELYNARHLCPHCEPOCPONGSYTCGFEPADOCVAACHYDXDP 600
Dy	LROGECEBEERVLQGLPRELYNARHLCPHCEPOCPONGSYTCGFEPADOCVAACHYDXDP 594
Db	LROGECEBEERVLQGLPRELYNARHLCPHCEPOCPONGSYTCGFEPADOCVAACHYDXDP 594
Dy	FCVARCPBGVYPDLSTWPIWKPPDEEGACOPCPINCTHSCVDLLDKGPCRAEPASPLTSI 660
Db	FCVARCPBGVYPDLSTWPIWKPPDEEGACOPCPINCTHSCVDLLDKGPCRAEPASPLTSI 660
Dy	VSAVWGILLVVLVGVFGGIIKKRKOOKIKRYTMRBLLOETELVEPLTPSGAMNQOMRI 720
Db	VSAVWGILLVVLVGVFGGIIKKRKOOKIKRYTMRBLLOETELVEPLTPSGAMNQOMRI 720
Dy	LKETELRKVKVLGSGAFGVYKGIWIPDENNVKI PVAILKLRENTSPKANKEILDAAYYW 780
Db	LKETELRKVKVLGSGAFGVYKGIWIPDENNVKI PVAILKLRENTSPKANKEILDAAYYW 774

Qy	781	AGVGSFVYSRLIGICTLSTVQLVQMPYGCILDHVRENRGLSGQDILNNCMQIAKMS	840
Db	775	AGVGSFVYSRLIGICTLSTVQLVQMPYGCILDHVRENRGLSGQDILNNCMQIAKMS	834
Qy	841	YLEGVRLVHRDLARNLVYKSPNHVKITDFGLARLIDIDEHYADGGKVPITKMMALESI	900
Db	835	YLEGVRLVHRDLARNLVYKSPNHVKITDFGLARLIDIDEHYADGGKVPITKMMALESI	894
Qy	901	LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAPEI PDLLEKGERLPQPICITIDVMI	960
Db	895	LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAPEI PDLLEKGERLPQPICITIDVMI	954
Qy	961	MVKWMIDSECRPFRELVSFSKMAKDPORFVVIQNEIDLCPASPLDSTFYRSLIEDDM	1020
Db	955	MVKWMIDSECRPFRELVSFSKMAKDPORFVVIQNEIDLCPASPLDSTFYRSLIEDDM	1014
Qy	1021	GDLVDAEEYLVPOGFCPPDPAFGAGVNHHRSSSTRSGGDLTGLSESEERASPS	1080
Db	1015	GDLVDAEEYLVPOGFCPPDPAFGAGVNHHRSSSTRSGGDLTGLSESEERASPS	1074
Qy	1081	LAPSEGASDVDFDGLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYAAPLTCS	1140
Db	1075	LAPSEGASDVDFDGLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYAAPLTCS	1134
Qy	1141	POPEYVNOPIVRRPOPSPBREGPLPAARPAGATLERAKTLSFGKGVVNDVAFGGAENP	1200
Db	1135	POPEYVNOPIVRRPOPSPBREGPLPAARPAGATLERAKTLSFGKGVVNDVAFGGAENP	1194
Qy	1201	EYLTPGGAAPQHPHPAPFAPNLVYWDODPERGAPSTFGCTPTAENPEYVLGLDVP	1260
Db	1195	EYLTPGGAAPQHPHPAPFAPNLVYWDODPERGAPSTFGCTPTAENPEYVLGLDVP	1254
Qy	1261	V 1261	
Db	1255	V 1255	
RESULT 2			
AAE12130 ID AAE12130 standard; Protein; 1255 AA.			
AAE12130:			
AC	18-DEC-2001	(first entry)	
DT			
XX			
DE	Human tyrosine kinase-type receptor, HER-2.		
KW	Therapeutic compound; major histocompatibility complex; vaccine;		
KW	antigenic peptide; MHC; immunoregulatory; immune response; HER-2;		
KW	adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;		
KW	antigen presenting cell; human; tyrosine kinase-type receptor.		
OS	Homo sapiens.		
XX			
PH	Key		
FT	Location/Qualifiers		
FT	774..782		
XX	Region		
XX	/note= "Antigenic epitope"		
XX	MO200168677-A2.		
XX	20-SEP-2001.		
XX	16-MAR-2001; 2001MO-US40328.		
XX	16-MAR-2000; 2000US-0527487.		
XX	(GENZ) GENZYME CORP.		
XX	Nicotette CA;		
XX			
XX	WPI; 2001-616284/71.		
XX	N-PSDB; AAD19731.		
XX			

PT	Novel synthetic therapeutic compound for inducing immune response and									
PT	for use in adoptive immunotherapy, has enhanced binding to major									
PT	histocompatibility molecules and enhanced immunoregulatory properties									
PT	-									
XX	Claim 4; Page 63-67; 69pp; English.									
XX										
CC	The invention relates to synthetic therapeutic compounds (antigenic									
CC	peptides) with enhanced binding to major histocompatibility complex									
CC	(MHC) molecules and enhanced immunoregulatory properties relative									
CC	to their natural counterparts. Compounds of the invention are useful									
CC	for inducing an immune response in a subject and for use in adoptive									
CC	immunotherapy. They are useful as components of anti-cancer vaccines									
CC	and to expand immune effector cells that are specific for cancers									
CC	characterised by expression of the breast cancer antigen, HER-2.									
CC	Polynucleotides that encode peptides of the invention are useful as									
CC	hybridisation probes and as primers for the detection of genes of gene									
CC	transcripts that are expressed in antigen presenting cells (APCs), to									
CC	confirm transduction of polynucleotides into host cells. The present									
CC	sequence is human tyrosine kinase-type receptor, HER-2. Compounds									
CC	of the invention are designed based on the HER-2 antigenic peptide									
CC	(774-782).									
XX										
SQ	Sequence 1255 AA;									
	Query Match 96.9%; Score 6632; DB 22; Length 1255;									
	Best Local Similarity 97.2%; Pred. No. 0;									
	Matches 1226; Conservative 8; Mismatches 21; Indels 6; Gaps 1;									
Qy	1	MELALCRWGLLALPPGAASVQCTGDKMLRLPASETHLDMLRLHYOGCQVQGNL	60							
Db	1	MELALCRWGLLALPPGAASVQCTGDKMLRLPASETHLDMLRLHYOGCQVQGNL	60							
Qy	61	ELTYLPTNASLSFLDIOEVQGYVLIANQVROPLQRLIRVGTOLFEDNYALAVDNG	120							
Db	61	ELTYLPTNASLSFLDIOEVQGYVLIANQVROPLQRLIRVGTOLFEDNYALAVDNG	120							
Qy	121	DPLNNTPTVNGASPGRLRELQRLSLTLKGVLIQRNPOLCYODTILMKDIFHKNOLA	180							
Db	121	DPLNNTPTVNGASPGRLRELQRLSLTLKGVLIQRNPOLCYODTILMKDIFHKNOLA	180							
Qy	181	LTLIDTNRSRACHPCSPMKSGSRMGSESEDCOSLTRVCAGACRCKGPLPTDCCHOC	240							
Db	181	LTLIDTNRSRACHPCSPMKSGSRMGSESEDCOSLTRVCAGACRCKGPLPTDCCHOC	240							
Qy	241	AAGCTGPHSDCLALPHNSGICELHCPALVYNTDTFESMPNEGRYTFGASCVTACP	300							
Db	241	AAGCTGPHSDCLALPHNSGICELHCPALVYNTDTFESMPNEGRYTFGASCVTACP	300							
Qy	301	YNYLSTVGSCTLVCPHNOEVTAEQRECKSKPCARVCYGLCMOYIKANSKFITG	360							
Db	301	YNYLSTVGSCTLVCPHNOEVTAEQRECKSKPCARVCYGLCMOYIKANSKFITG	360							
Qy	361	ELBFGCKKIFGSLAFPLPESFGDPASNTAPLOEOLVFEITLTYLISAMPDLP	420							
Db	361	ELBFGCKKIFGSLAFPLPESFGDPASNTAPLOEOLVFEITLTYLISAMPDLP	420							
Qy	421	DLSVFQNLQVIRGRILHNGAVSLTLQIGISWLGRLSRLREGSGIALIHNTHLCFVATV	480							
Db	421	DLSVFQNLQVIRGRILHNGAVSLTLQIGISWLGRLSRLREGSGIALIHNTHLCFVATV	480							
Qy	481	PMDQLFRFNNFTYSFMLRVKVSASHLECYGBGLACHOLCARGHCMGPGFTQCVNSQF	540							
Db	481	PMDQLFRFNNFTYSFMLRVKVSASHLECYGBGLACHOLCARGHCMGPGFTQCVNSQF	534							
Qy	541	LRGQECVBEGRVLOGLPREYVNAHCLPCHECOPONGSVTCFPEADQCVACHYXDP	600							
Db	535	LRGQECVBEGRVLOGLPREYVNAHCLPCHECOPONGSVTCFPEADQCVACHYXDP	594							
Qy	601	FCVACPSGVKPDLSYMPINKFPDEBGAQPCPINCTHSCVDLDDKCCPAEQRASPLTSI	660							
Db	595	FCVACPSGVKPDLSYMPINKFPDEBGAQPCPINCTHSCVDLDDKCCPAEQRASPLTSI	654							

```
QY 661 VSAVVGILLVVLGVFGILLIKRRQOKIRKYMTRRLLOETELVEPLTPSGAMPNOQMRI 720
DB 655 VSAVVGILLVVLGVFGILLIKRRQOKIRKYMTRRLLOETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPGENVKIKVAILKVBRENTSPKANKEILDEAYVM 780
DB 715 LKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVAILKVBRENTSPKANKEILDEAYVM 774
QY 781 AGVSPYVSRLLGICLTSTVOLVTOLMPYGCILDHVRENRRGLSGODLLNMCQIAKMS 840
DB 775 AGVSPYVSRLLGICLTSTVOLVTOLMPYGCILDHVRENRRGLSGODLLNMCQIAKMS 834
QY 841 YLEDEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPKIMWALESI 900
DB 835 YLEDEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPKIMWALESI 894
QY 901 LRRFTTHOSDVWSYGVYTWELMTFGAKPYDGIIPAEIIDLLEKGRLLPOPPICITIDVYMI 960
DB 895 LRRFTTHOSDVWSYGVYTWELMTFGAKPYDGIIPAEIIDLLEKGRLLPOPPICITIDVYMI 954
QY 961 MVKCMIMIDSECRPFRELVSFSSRMARDPQRFVVIQNEIDLGPASPPLDSTFYRSLLEDDDM 1020
DB 955 MVKCMIMIDSECRPFRELVSFSSRMARDPQRFVVIQNEIDLGPASPPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEELVLPQCGFCPCDPAPGAGVNHHRHSSSTRSGGDLTTLGLEPSEBEAPRSP 1080
DB 1015 GDLVDAEELVLPQCGFCPCDPAPGAGVNHHRHSSSTRSGGDLTTLGLEPSEBEAPRSP 1074
QY 1081 LAPSEAGSDVFDLGLGMAKGLQSLPTHPSPLORESPTVPLBSETGCVAPLTCSS 1140
DB 1075 LAPSEAGSDVFDLGLGMAKGLQSLPTHPSPLORESPTVPLBSETGCVAPLTCSS 1134
QY 1141 POPPYVNOPDVRPOPSPREGPLPAARPAATLERAKTLSPGKGVNVDPVAFGAVENP 1200
DB 1135 POPPYVNOPDVRPOPSPREGPLPAARPAATLERAKTLSPGKGVNVDPVAFGAVENP 1194
QY 1201 EYLTPQCGAAPQHPHPAFSPAFDNLVYWDODPEERGAPOSTFKGTPTAENPEYLGIDVP 1260
DB 1195 EYLTPQCGAAPQHPHPAFSPAFDNLVYWDODPEERGAPOSTFKGTPTAENPEYLGIDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 3
AAB60167
ID AAB60167 standard; Proteoin; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KM Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX
OS Homo sapiens.
XX Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX
PA 16-MAR-2000; 2000US-0189844.
XX (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R,
```

```
XX WP1; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6632; DB 22; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1226; Conservative 8; Mismatches 21; Indels 6; Gaps 1;
QY 1 MELALCRMGILLALLPPGAASTOYCTGTDMLRLPASPETHLDMRLHYOCQVVOGNTL 60
DB 1 MELALCRMGILLALLPPGAASTOYCTGTDMLRLPASPETHLDMRLHYOCQVVOGNTL 60
QY 61 ELTYLPTNASISFLDIOEVGVYLIANOVROYVLORLRIVRGTOLEFEDNYALAVLNG 120
DB 61 ELTYLPTNASISFLDIOEVGVYLIANOVROYVLORLRIVRGTOLEFEDNYALAVLNG 120
QY 121 DPLNNTPTVGTASPGGLRELOLRSLTELKGVLIQRRPOLCYODTILMKOIFHONOLA 180
DB 121 DPLNNTPTVGTASPGGLRELOLRSLTELKGVLIQRRPOLCYODTILMKOIFHONOLA 180
QY 181 LTLIDITNRSRACHPCSPKCKSRGCESSSDCOSLTRVCAGGACRCKGPIPTDCHEOC 240
DB 181 LTLIDITNRSRACHPCSPKCKSRGCESSSDCOSLTRVCAGGACRCKGPIPTDCHEOC 240
QY 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVYNTDTFESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVYNTDTFESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNDEVTADGTORCEKCSKPARVCYIGMOTYIANKSFIGIT 360
DB 301 YNYLSTDVGSCTLVCPHLNDEVTADGTORCEKCSKPARVCYIGMOTYIANKSFIGIT 360
QY 361 ELSEFAGCKIIFGSLAFLPESFDPGDPASNTAPLOPEOLOVFETLEETGYLISAMPDSL 420
DB 361 IGFPAAGCKIIFGSLAFLPESFDPGDPASNTAPLOPEOLOVFETLEETGYLISAMPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWLGILRSLEISGIALIHNHTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWLGILRSLEISGIALIHNHTHLCFVHTV 480
QY 481 PMDQLFFRNFTVFSFHLRVPVVSASHLECYGEGLAGHQLCARHGCMWPGGTQCVNCGOF 540
DB 481 PMDQLFFRNFTVFSFHLRVPVVSASHLECYGEGLAGHQLCARHGCMWPGGTQCVNCGOF 534
QY 541 LRGECEVCEBRCVLOGLPREYVNAHRCPLCPHECOPONGSVTCFPGPADQCAACHYXDP 600
DB 541 LRGECEVCEBRCVLOGLPREYVNAHRCPLCPHECOPONGSVTCFPGPADQCAACHYXDP 594
QY 601 FCVARCPGKVPDLSYMPIMKFPDEBAGCPPICTHSCVYDLDDKCPAPQASPLTSI 660
DB 595 FCVARCPGKVPDLSYMPIMKFPDEBAGCPPICTHSCVYDLDDKCPAPQASPLTSI 654
QY 661 VSAVVGILLVVLGVFGILLIKRRQOKIRKYMTRRLLOETELVEPLTPSGAMPNOQMRI 720
DB 655 VSAVVGILLVVLGVFGILLIKRRQOKIRKYMTRRLLOETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVAILKVBRENTSPKANKEILDEAYVM 780
```

```

Db      715 LKETELKVKVILSGAGCTVYKGIWIPGENVKI PVAIKVLRNTSPANKELIDEAVM 774
Qy      781 AGVGSFYVSRLLGICLTSTVOLVTOLMFYGCLLDHVRNRRGLSGODLLNMCQIAKMS 840
Db      775 AGVGSFYVSRLLGICLTSTVOLVTOLMFYGCLLDHVRNRRGLSGODLLNMCQIAKMS 834
Qy      841 YLEEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETBYHADGSKVPIKMMALESI 900
Db      835 YLEEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETBYHADGSKVPIKMMALESI 894
Qy      901 LRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAEITDLEKGRLLGOPICITDVYMI 960
Db      895 LRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAEITDLEKGRLLGOPICITDVYMI 954
Qy      961 MVKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQNDLGPASPLDSTFYRSLLEDDM 1020
Db      955 MVKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQNDLGPASPLDSTFYRSLLEDDM 1014
Qy      1021 GDLYDAEEYLVPOCGFFCPDPAPGAGVHHRHSSSTRSGGDLTLGLPSEEEAPRSP 1080
Db      1015 GDLYDAEEYLVPOCGFFCPDPAPGAGVHHRHSSSTRSGGDLTLGLPSEEEAPRSP 1074
Qy      1081 LAPESGSDVFDGDLGMAKGLQSLPTHDPSPLOKRSSEDPVPLPSETDGYVAPLTCG 1140
Db      1075 LAPESGSDVFDGDLGMAKGLQSLPTHDPSPLOKRSSEDPVPLPSETDGYVAPLTCG 1134
Qy      1141 POPEYVQPDVPPQPSPREGEPPLPARPAGATLEBAKTLSPGKGVNDVAFGGAVENP 1200
Db      1135 POPEYVQPDVPPQPSPREGEPPLPARPAGATLEBAKTLSPGKGVNDVAFGGAVENP 1194
Qy      1201 EYLTPQGAAPQPPPPAPFADNLVYWDODPPERGAPESTFKGTPFAENPEYIGLDVP 1260
Db      1195 EYLTPQGAAPQPPPPAPFADNLVYWDODPPERGAPESTFKGTPFAENPEYIGLDVP 1254
Qy      1261 V 1261
Db      1255 V 1255

```

RESULT 4
AAU74545
ID AAU74545 standard; Protein, 1255 AA.

AAU74545;
23-APR-2002 (first entry)
Human HER2 (ErbB2) polypeptide.

Human; HER2; ErbB; epidermal growth factor receptor; receptor;
anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
glial disorder; astrocytal disorder; hypothalamic disorder;
glandular disorder; macropneagal disorder; epithelial disorder;
stromal disorder; blastocoele disorder; inflammatory disorder;
angiogenic disorder; immunological disorder.

Homo sapiens.

US2002001587-A1.

03-JAN-2002.

16-MAR-2001; 2001US-0811123.

16-MAR-2000; 2000US-189844P.

05-OCT-2000; 2000US-238327P.

ERIC/ ERICKSON S.
PA (SCHN/ SCHWALL R.
PA (SLIW/ SLIKOWSKI M.

```

PI      Erickson S, Schwall R, Slikowski M,
XX      WPI, 2002-163686/21.
DR      N-PSDB; ABR14058.
XX      Treating tumour characterised by overexpression of epidermal growth
PT      factor receptor, ErbB or cancer in mammal, comprises administering
PT      anti-ErbB antibody-maytansinoid conjugate to the mammal
XX      Example 3; Fig 7; 93pp; English.
PS
CC      The invention relates to treating a tumour in a mammal, where the tumour
CC      is characterised by the overexpression of an epidermal growth factor
CC      receptor (ErbB) and does not respond or responds poorly, to treatment
CC      with an anti-ErbB antibody, comprising administering to the mammal an
CC      anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC      treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC      salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC      prostate and bladder, preferably breast cancer. The breast cancer is a
CC      metastatic breast cancer or an aggressive form of metastatic breast
CC      cancer which overexpresses ErbB2. The method is also useful for treating
CC      neuronal, glial, astrocytal, hypothalamic, glandular, macropneagal,
CC      epithelial, stromal, blastocoele, inflammatory, angiogenic and
CC      immunological disorders. This sequence represents the human HER2 (ErbB2)
CC      polypeptide of the invention.
XX
SQ      Sequence 1255 AA;
XX
Query Match      96.9%; Score 6632; DB 23; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1226; Conservative 8; Mismatches 21; Indels 6; Gaps 1;

```

```

Qy      1 MEIALACRWGLLALLPFGAASVQCTGDMKRLPASPETHLMDLRHLYOCQVQVGL 60
Db      1 MEIALACRWGLLALLPFGAASVQCTGDMKRLPASPETHLMDLRHLYOCQVQVGL 60
Qy      61 ELTYLPTNASLFLQDIOEVQYVLIANOVROVPLQRLRYRGTQLFEDNYALVLDNG 120
Db      61 ELTYLPTNASLFLQDIOEVQYVLIANOVROVPLQRLRYRGTQLFEDNYALVLDNG 120
Qy      121 DPLNNTPTVTGASPGGLRELOLRSTELTKGVLQIRNPOLCYDTIIMKDI FHKNOIA 180
Db      121 DPLNNTPTVTGASPGGLRELOLRSTELTKGVLQIRNPOLCYDTIIMKDI FHKNOIA 180
Qy      181 LTLIDNRSRACHPCSPWCKSGRCWGESSEDCQSLTRVCAAGCARCGPLPTDCHEQC 240
Db      181 LTLIDNRSRACHPCSPWCKSGRCWGESSEDCQSLTRVCAAGCARCGPLPTDCHEQC 240
Qy      241 AAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDTFESMPNREGRTFGASCVTACP 300
Db      241 AAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDTFESMPNREGRTFGASCVTACP 300
Qy      301 YNVLSTVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNVLSTVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
Qy      361 ELERAGCKKIFGSLAFIPESFGDPASNTAPLOPQLOVFETLEITGYLIASAMPDISLP 420
Db      361 ELERAGCKKIFGSLAFIPESFGDPASNTAPLOPQLOVFETLEITGYLIASAMPDISLP 420
Qy      421 DLSTVFQNLQVIRGRIILNHGAYSLTLOGIGISMLGIRSLREISGSLALHHNTHLCFVHTV 480
Db      421 DLSTVFQNLQVIRGRIILNHGAYSLTLOGIGISMLGIRSLREISGSLALHHNTHLCFVHTV 480
Qy      481 PMDQLFRNNFTVSFWLVRPVYASHTLEECVGEGLACHQLCARGHCWPGPOTCVNCSOF 540
Db      481 PMDQLFRNNFTVSFWLVRPVYASHTLEECVGEGLACHQLCARGHCWPGPOTCVNCSOF 540
Qy      541 LRGECEVECRVLOGLFREYVNAHCLPCHECQPOGNSVTTCFPEADQCAACHYKXP 600
Db      541 LRGECEVECRVLOGLFREYVNAHCLPCHECQPOGNSVTTCFPEADQCAACHYKXP 600
Qy      594 FCVARCPGKVPDLSTYPIWKFPEDEGACQCPINCHSGCVTDLDKCGAPQASPLTISI 660
Db      594 FCVARCPGKVPDLSTYPIWKFPEDEGACQCPINCHSGCVTDLDKCGAPQASPLTISI 660

```

```
DB 595 FCVACPCPSGVKSDLSPMTWKPPDEGACOPCPINCTHSCVLLDDKGCPCAEQASPLTSI 654
QY 661 VSAVVGILLVVLGVVFGILIKRROOKIRKYMTRRLLOETELVEPIITPGAMPNOQM1 720
DB 655 VSAVVGILLVVLGVVFGILIKRROOKIRKYMTRRLLOETELVEPIITPGAMPNOQM1 714
QY 721 LKETELRKVKVLGSGAFGTVMKGIWIPDGENVKIPVAIKVLENTSPKANKETLDEAYM 780
DB 715 LKETELRKVKVLGSGAFGTVMKGIWIPDGENVKIPVAIKVLENTSPKANKETLDEAYM 774
QY 761 AGVGSPPVSRLLIGICTSTVOLVTQMLPYGCLLDHVRNRRGLSGODLLNMCMQIAKMS 840
DB 775 AGVGSPPVSRLLIGICTSTVOLVTQMLPYGCLLDHVRNRRGLSGODLLNMCMQIAKMS 834
QY 841 YLEDVRLVHRDLAARVNLKSPNHVKITDFGLARLLDIDETEHNAQGVPIKMMALESI 900
DB 835 YLEDVRLVHRDLAARVNLKSPNHVKITDFGLARLLDIDETEHNAQGVPIKMMALESI 894
QY 901 LRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPOPICTIDVYMI 960
DB 895 LRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPOPICTIDVYMI 954
QY 961 MYKCMWIDSECRPRRELVSERSRMARDPQRVYVIONEDLGPASPLDSTFYSLLEDDM 1020
DB 955 MYKCMWIDSECRPRRELVSERSRMARDPQRVYVIONEDLGPASPLDSTFYSLLEDDM 1014
QY 1021 GDVLVAEEYLPVQOGFPCDPAPAGAGVHHRRSSSTSGGDLTLGLEPSEEPSP 1080
DB 1015 GDVLVAEEYLPVQOGFPCDPAPAGAGVHHRRSSSTSGGDLTLGLEPSEEPSP 1074
QY 1081 LAPSEGASDVDFDGLGMAAGLQSLPTHDSPLQRYSEDPVPLPSETDGVAPLTS 1140
DB 1075 LAPSEGASDVDFDGLGMAAGLQSLPTHDSPLQRYSEDPVPLPSETDGVAPLTS 1134
QY 1141 POPEYVNOQDVAPPOPPSPREGPLPAARPGATLERAKTILSPGKNGVYKOVFAFGAVENP 1200
DB 1135 POPEYVNOQDVAPPOPPSPREGPLPAARPGATLERAKTILSPGKNGVYKOVFAFGAVENP 1194
QY 1201 EYLTPQGGAPOPPHPPAPAFNDLYYWDODPPERGAPSTPFKGTPTAENEYVGLDVP 1260
DB 1195 EYLTPQGGAPOPPHPPAPAFNDLYYWDODPPERGAPSTPFKGTPTAENEYVGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu: c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 676..1255
FT Domain
FT /label= Intracellular_domain
FT /note= "claimed domain, useful for immunisation"
XX
XX MO9630514-A1.
XX 03-OCT-1996.
XX
```

```
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNITV ) UNITV WASHINGTON.
XX
PI Cheever MA, Distle ML;
XX
DR MPI: 1996-455361/45.
DR N-PSDB; AAT40739.
XX
XX
PT DNA encoding HER-2-neu polypeptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2, Page 56-61; 71pp; English.
XX
CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA:
XX
Query Match 96.8%; Score 6626; DB 17; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
XX
QY 1 MELALACRWGLLALLPGCASTOYCTGDMKRLRPASPTHLDMLRHLYOGCQVQGNL 60
DB 1 MELALACRWGLLALLPGCASTOYCTGDMKRLRPASPTHLDMLRHLYOGCQVQGNL 60
QY 61 ELTYLPTNASSLFLODIOVGQYVLIANQVAVQPLQRIYRGQLFEDNALAVLDNG 120
DB 61 ELTYLPTNASSLFLODIOVGQYVLIANQVAVQPLQRIYRGQLFEDNALAVLDNG 120
QY 121 DELNNTTPYTGASPGGLRELQLRSLEILKGVLIQRNPQLCYOPIILMKDIFHKNNQLA 180
DB 121 DELNNTTPYTGASPGGLRELQLRSLEILKGVLIQRNPQLCYOPIILMKDIFHKNNQLA 180
QY 181 LTLIDTNSRACHPCSPMKCKSGRCESESDQSILTRYCAGGCARCKPLPTDCHEQC 240
DB 181 LTLIDTNSRACHPCSPMKCKSGRCESESDQSILTRYCAGGCARCKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLALHFHSGICELHCPALVTYNTDFESMPNBEGRYTFGASCYACP 300
DB 241 AAGCTGPKHSDCLALHFHSGICELHCPALVTYNTDFESMPNBEGRYTFGASCYACP 300
QY 301 YNYLSTDVSGCTLVCPILHNOEYTAEDGTORCECKSPCARVCYGLQMOYIKANSKPIGIT 360
DB 301 YNYLSTDVSGCTLVCPILHNOEYTAEDGTORCECKSPCARVCYGLQMOYIKANSKPIGIT 360
QY 361 ELIEFAGCKKIRGSALFLPESPDGPASNTAPLQEPOLQVFEETLEITGYLYISAMPDSIP 420
DB 361 IOEFAGCKKIRGSALFLPESPDGPASNTAPLQEPOLQVFEETLEITGYLYISAMPDSIP 420
QY 421 DLSVFQNTQVIRGRILHNGAYSLLTQIGISMLGIRSLREIGSGALILHNNHLLCPVHV 480
DB 421 DLSVFQNTQVIRGRILHNGAYSLLTQIGISMLGIRSLREIGSGALILHNNHLLCPVHV 480
QY 481 FMDQLFRNNFTVSMFLRVPKVASHLSECVGEGLAGHQLCARGHCKWGPGPTQVCNCSQF 540
DB 481 FMDQLFRNNFTVSMFLRVPKVASHLSECVGEGLAGHQLCARGHCKWGPGPTQVCNCSQF 540
QY 541 LRGECEVECKVLOGLPREYVNAARCLPCHPECOPQNSVTCFGEADQCAAHYKDP 600
DB 541 LRGECEVECKVLOGLPREYVNAARCLPCHPECOPQNSVTCFGEADQCAAHYKDP 600
QY 535 LRGECEVECKVLOGLPREYVNAARCLPCHPECOPQNSVTCFGEADQCAAHYKDP 594
DB 535 LRGECEVECKVLOGLPREYVNAARCLPCHPECOPQNSVTCFGEADQCAAHYKDP 594
```

```

OY 601 FCVACPSGVKPDLSYMPIWKFPEDEGACOPCPINCTHS CVLDLDDKCAEORASPLTSTI 660
DB 595 FCVACPSGVKPDLSYMPIWKFPEDEGACOPCPINCTHS CVLDLDDKCAEORASPLTSTI 654
OY 661 USAVVGILLVVLGVVFGILIKRROOKIKRYTMRRLQETELVEPLTPSGAMPNOQOMI 720
DB 655 ISAVVGILLVVLGVVFGILIKRROOKIKRYTMRRLQETELVEPLTPSGAMPNOQOMI 714
OY 721 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKELIDEAAYM 780
DB 715 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKELIDEAAYM 774
OY 761 AGVSPVYSRLIGLICTSTVOLVTOUMPGCLLDHRENRGRGLSGDILLNMCQIACKMS 840
DB 775 AGVSPVYSRLIGLICTSTVOLVTOUMPGCLLDHRENRGRGLSGDILLNMCQIACKMS 834
OY 841 YLEDRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKIMMALESI 900
DB 835 YLEDRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKIMMALESI 894
OY 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
DB 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
OY 961 MWKCMIDSECRPRELVESESRMARDBORFVVIQNEDELGPASPLDSTFFYSLLDDDM 1020
DB 955 MWKCMIDSECRPRELVESESRMARDBORFVVIQNEDELGPASPLDSTFFYSLLDDDM 1014
OY 1021 GDLVDAEELVPOQGFPCDPAPAGAGMWHRRSSSTSSGGDDLTLGLEPSEBEAPRSP 1080
DB 1015 GDLVDAEELVPOQGFPCDPAPAGAGMWHRRSSSTSSGGDDLTLGLEPSEBEAPRSP 1074
OY 1081 LAPSEGASDVFDDGLGWAAGKGLSLPTHDPSPLORYSEDPVPLPSETIDGVYAPLTCS 1140
DB 1075 LAPSEGASDVFDDGLGWAAGKGLSLPTHDPSPLORYSEDPVPLPSETIDGVYAPLTCS 1134
OY 1141 POPEVYNOPVYRPOPSPREBPLPAARPAKATLEBAKTLSPKNGVYKOVFAFGAVENP 1200
DB 1135 POPEVYNOPVYRPOPSPREBPLPAARPAKATLEBAKTLSPKNGVYKOVFAFGAVENP 1194
OY 1201 EYLTPQGAAPQHPHPPAPAFNDLYYWDQDPERGAPSTFGKTPETAENEYGLDVP 1260
DB 1195 EYLTPQGAAPQHPHPPAPAFNDLYYWDQDPERGAPSTFGKTPETAENEYGLDVP 1254
OY 1261 V 1261
DB 1255 V 1255

```

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.

```

AC AAW92406;
DT 21-APR-1999 (first entry)
DE Human HER-2/neu oncogene protein.
EE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KM malignancy; treatment; tumour.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Region 676..1255
FT /note="region which elicits immune response"
XX
XX US5869445-A.
XX 09-FEB-1999.
XX 01-APR-1996; 96US-0625101.

```

```

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX (UNIT ) UNIV WASHINGTON.
XX Cheever MA, Distle ML;
XX MPI: 1999-152835/13.
DR N-PSDB; AAX01912.
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
XX Claim 3; Column 31-38; 26pp; English.
XX
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
XX of this protein is used in a method for eliciting or enhancing an immune
XX response to HER-2/neu protein. The polypeptide can stimulate T cells and
XX B cells to produce an immune response to the HER-2/neu protein. The
XX method can be used for immunisation against a malignancy in which the
XX HER-2/neu oncogene is associated and in the treatment of an existing
XX tumour, or to prevent tumour occurrence or recurrence.
SQ Sequence 1255 AA;
Query Match 96.8%; Score 6626; DB 20; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
OY 1 METALACRMGILLALIPGASTOYCTGDMKRLPASPETHLMDRLHYOGGVQGNL 60
DB 1 METALACRMGILLALIPGASTOYCTGDMKRLPASPETHLMDRLHYOGGVQGNL 60
OY 61 ELTYLPTNASLSFLDDIOEVGYVLIANQVQVPLQRLIRYRGTLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDDIOEVGYVLIANQVQVPLQRLIRYRGTLFEDNYALAVLDNG 120
OY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYODTILMKDIFHXNOLA 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYODTILMKDIFHXNOLA 180
OY 181 LTLIDNRSRACHPCSPCKSGRCSGSESDQSILRTVCAGGACRCKPLPTDCHEOC 240
DB 181 LTLIDNRSRACHPCSPCKSGRCSGSESDQSILRTVCAGGACRCKPLPTDCHEOC 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFCASCTYACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFCASCTYACP 300
OY 301 YNYLSTDVGSCTLVCPPLHNOEVTABDGTORCEKSPCARVCYGLGMOYIKANSKFTIGIT 360
DB 301 YNYLSTDVGSCTLVCPPLHNOEVTABDGTORCEKSPCARVCYGLGMEHLREBRAVTSAN 360
OY 361 ELEFAGCKKIRGSLAFPLESPFDGPASTNAPLQEPOLQVFEFTLEITGLYLIISAMPDCLP 420
DB 361 IOEFAGCKKIRGSLAFPLESPFDGPASTNAPLQEPOLQVFEFTLEITGLYLIISAMPDCLP 420
OY 421 DLSVFQNTQVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELGSLALHNNHNLGCFVHTV 480
DB 421 DLSVFQNTQVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELGSLALHNNHNLGCFVHTV 480
OY 481 PMDOLFRPNNTVTSFWLAVPVKVASHLKECVGEGLAGLQCLARGHCHWGPGPTQVCNCSQF 540
DB 481 PMDOLFRPNNTVTSFWLAVPVKVASHLKECVGEGLAGLQCLARGHCHWGPGPTQVCNCSQF 534
OY 541 LRGOECVEECVLOGLPREYVNAHCHLPCHRECOPONSVCYCFGEAQQCAVCAHYKXPP 600
DB 541 LRGOECVEECVLOGLPREYVNAHCHLPCHRECOPONSVCYCFGEAQQCAVCAHYKXPP 594

```



```

QY 601 FCVARCPSGVKBDLSYMPIWKPEDEBACQPCPINCSTHSCVDLDDKGCRAEQRASPLTISI
DB 595 FCVARCPSGVKBDLSYMPIWKPEDEBACQPCPINCSTHSCVDLDDKGCRAEQRASPLTISI
QY 661 VSAVVGILLVVLGVVFGILLIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRI
DB 655 ISAVVGILLVVLGVVFGILLIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRI
QY 721 LKETELRKVKVLSGSAFGTVYKGIWI PDENVKI PAIVKYLRENTSPKXNKETLDDAAYM
DB 715 LKETELRKVKVLSGSAFGTVYKGIWI PDENVKI PAIVKYLRENTSPKXNKETLDDAAYM
QY 781 AGVGSFYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENGRGLSGODLNMCMQIAKGS
DB 775 AGVGSFYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENGRGLSGODLNMCMQIAKGS
QY 841 YLEDVRLVHRDLAARVNLVKSPPNHVKITDPTGLARLLDIDETEHADGKVPKIMWALESI
DB 835 YLEDVRLVHRDLAARVNLVKSPPNHVKITDPTGLARLLDIDETEHADGKVPKIMWALESI
QY 901 LRRRFTHSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPOPTCTIDVMI
DB 895 LRRRFTHSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPOPTCTIDVMI
QY 961 MWKCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM
DB 955 MWKCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM
QY 1021 GLVDAEYLVLPQOGFCDDPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEEARSP
DB 1015 GLVDAEYLVLPQOGFCDDPAPAGAGMHHRRSSSTSGGDDLTLGLEPSEEARSP
QY 1081 LAPSEGASDVDFDGLGMAKAGLOSLPTHDPSPPLRYSSEDPVPLPSETDGVVALTCS
DB 1075 LAPSEGASDVDFDGLGMAKAGLOSLPTHDPSPPLRYSSEDPVPLPSETDGVVALTCS
QY 1141 PPEYVNPQDVP RPQPSREGPLPAARPAATLERAKTSLPCQNGVVKDVFAGAVENP
DB 1135 PPEYVNPQDVP RPQPSREGPLPAARPAATLERAKTSLPCQNGVVKDVFAGAVENP
QY 1201 EYLTPOGGAAPQHPAPASPAFNDLYYNDOPPEPGAPSPSTFKGPTAENPEYLGLDVP
DB 1195 EYLTPOGGAAPQHPAPASPAFNDLYYNDOPPEPGAPSPSTFKGPTAENPEYLGLDVP
QY 1261 V 1261
DB 1255 V 1255

```

RESULT 7
AAB21198
ID AAB21198 standard: protein; 1255 AA.

```

XX AAB21198;
AC 12-JAN-2001 (first entry)
DT 12-JAN-2001 (first entry)
DE Human HER-2/neu protein.
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX colon cancer.
XX Homo sapiens.
XX OS
XX WO200044899-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000WO-US02164.
XX 29-JAN-1999; 99US-0117976.
XX

```

```

PA (CORI-) CORIXA CORP.
PA (SMIK ) SMITHKLINE BEECHAM.
XX Cheever MA, Gheyssen D;
XX WPI: 2000-505976/45.
DR N-PSDB: AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
CC
SQ Sequence 1255 AA;

```

Query Match 96.8%; Score 6626; DB 21; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;

```

QY 1 MELALCRGGLLALLPPGAASVCTGTDMLRLPASFEHLDMRLHYOGGVQGNL 60
DB 1 MELALCRGGLLALLPPGAASVCTGTDMLRLPASFEHLDMRLHYOGGVQGNL 60
QY 61 EYLTPTNASLFLDIOGVGVVLIANQVQVPLQRIYRGTOLEFDNVALAVLDNG 120
DB 61 EYLTPTNASLFLDIOGVGVVLIANQVQVPLQRIYRGTOLEFDNVALAVLDNG 120
QY 121 DPLNNTPTVYTGASPGGLRELQRLSTLEILKGVLIQRPOLCYODTILKMDIFHKNNOLA 180
DB 121 DPLNNTPTVYTGASPGGLRELQRLSTLEILKGVLIQRPOLCYODTILKMDIFHKNNOLA 180
QY 181 LTLIDTNRBACHPGSPMKSGRCWGESSEDCQSLTRYTCAGGCARCKRPLPDDCCEQC 240
DB 181 LTLIDTNRBACHPGSPMKSGRCWGESSEDCQSLTRYTCAGGCARCKRPLPDDCCEQC 240
QY 241 AAGCTGPKHSDCLACHFPHSGICELHCPALVTYNTDFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFPHSGICELHCPALVTYNTDFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPLEHNOEYTABDGTQRCCKSPKCARVCGYGLGMQYIKANSKEFIGIT 360
DB 301 YNYLSTDVSGCTLVCPLEHNOEYTABDGTQRCCKSPKCARVCGYGLGMQYIKANSKEFIGIT 360
QY 361 ELEPACCKKIFGSLAFLEPSFPGDPASTNAPLOPBOLOVFETLEBITGYLYISAMPDSL 420
DB 361 ELEPACCKKIFGSLAFLEPSFPGDPASTNAPLOPBOLOVFETLEBITGYLYISAMPDSL 420
QY 421 DLSTVONLOVIRGLIHNGAYSLTLOGLISWLGRLSRLBELSGALIHNNHILCFVHV 480
DB 421 DLSTVONLOVIRGLIHNGAYSLTLOGLISWLGRLSRLBELSGALIHNNHILCFVHV 480
QY 481 PWDOLFRRNFTVSFMLRVKVSASHLECEVGEGLACHQLCARGHCWGPGPTQCVVCSOF 540
DB 481 PWDOLFRRNFTVSFMLRVKVSASHLECEVGEGLACHQLCARGHCWGPGPTQCVVCSOF 540
QY 541 LRGOECVECRVLOGLPREYVNAARCLPHEPCQONKSVYTFGPBADCVCACAHYKDP 600
DB 541 LRGOECVECRVLOGLPREYVNAARCLPHEPCQONKSVYTFGPBADCVCACAHYKDP 600
QY 594 LGQECVECRVLOGLPREYVNAARCLPHEPCQONKSVYTFGPBADCVCACAHYKDP 594
DB 594 LGQECVECRVLOGLPREYVNAARCLPHEPCQONKSVYTFGPBADCVCACAHYKDP 594
QY 601 FCVARCPSGVKBDLSYMPIWKPEDEBACQPCPINCSTHSCVDLDDKGCRAEQRASPLTISI 660
DB 601 FCVARCPSGVKBDLSYMPIWKPEDEBACQPCPINCSTHSCVDLDDKGCRAEQRASPLTISI 660

```

```

Db      |||||
595 FCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCHTSCVDLDDKCGPAERASPLTSI 654
Qy      |||||
661 USAVVGILLVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAMPNOAKRI 720
Db      |||||
655 ISAVVGIILVVLGVVFGILLKROOKIRKTYMRRLQETELVEPLTPSGAMPNOAKRI 714
Qy      |||||
721 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDEAYVM 780
Db      |||||
715 LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDEAYVM 774
Qy      |||||
781 AGVSPYVSRLLGICLTSTVOLVTOLMPYGLLDHVRENRGLSGODLNMCMQIAKWS 840
Db      |||||
775 AGVSPYVSRLLGICLTSTVOLVTOLMPYGLLDHVRENRGLSGODLNMCMQIAKWS 834
Qy      |||||
841 YLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGGKVPKMMALSI 900
Db      |||||
835 YLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGGKVPKMMALSI 894
Qy      |||||
901 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYMI 960
Db      |||||
895 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPICTIDVYMI 954
Qy      |||||
961 MVKCMIDSECRPRRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1020
Db      |||||
955 MVKCMIDSECRPRRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1014
Qy      |||||
1021 GDLYDAEEYLPQOGFCPPDPAPGAGVNHHRSSSTRSGGDLTLGLEPSEEPAPSP 1080
Db      |||||
1015 GDLYDAEEYLPQOGFCPPDPAPGAGVNHHRSSSTRSGGDLTLGLEPSEEPAPSP 1074
Qy      |||||
1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKRYSEDPVLPSETDGYVAPLCS 1140
Db      |||||
1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKRYSEDPVLPSETDGYVAPLCS 1134
Qy      |||||
1141 POPRYVNPQVPRPQPSRREGPLPAARAGATLERAKTLSKKNVYQDVFAFGAVENP 1200
Db      |||||
1135 POPRYVNPQVPRPQPSRREGPLPAARAGATLERAKTLSKKNVYQDVFAFGAVENP 1194
Qy      |||||
1201 EYLTPOGGAAPQPPAPFSPAFDLYYWDODPPERGA PPTFFKTPTAENEYVGLDVP 1260
Db      |||||
1195 EYLTPOGGAAPQPPAPFSPAFDLYYWDODPPERGA PPTFFKTPTAENEYVGLDVP 1254
Qy      |||||
1261 V 1261
Db      |||||
1255 V 1255

```

RESULT 8
AA184780
ID AA184780 standard; Protein, 1255 AA.

AA184780;
08-AUG-2000 (first entry)
Amino acid sequence of the *SPICE* *erbB-2* receptor protein.
SPICE *erbB-2* receptor protein; cell transformation disorder; cancer;
tumor cell proliferation; tissue degeneration; arthropathy;
bone resorption; inflammatory disease; degenerative disorder;
wound healing.
Homo sapiens.
WO200020579-A1.
13-APR-2000.
01-OCT-1999; 99WO-CA00912.
02-OCT-1998; 98US-0165192.

```

PA      (UNYC-) UNIV MCMMASTER.
XX      Muller WJ, Siegel PM;
PI      WPI, 2000-303768/26.
DR      N-PSDB; AAA14812.
XX      Nucleic acid encoding an erbB-2 receptor protein designated SPICE
PT      erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX      Claim 3; Fig 2; 60pp; English.
PS
CC      The present sequence represents a SPICE erbB-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erbB-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPICE erbB-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPICE erbB-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPICE erbB-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
SQ      Sequence 1255 AA;
Query Match 96.8%; Score 6626; DB 21; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
Qy      1 MEIALACWGLLALLPFGASTQVCTGDKMLRPAPEHLLMLRLHYOCQVQVGNL 60
Db      1 MEIALACWGLLALLPFGASTQVCTGDKMLRPAPEHLLMLRLHYOCQVQVGNL 60
Qy      61 ELTYLPTNASLFLQDIOEVGYVLIANOVROYPLQRLRIYRGTQLFEDNALAVLNG 120
Db      61 ELTYLPTNASLFLQDIOEVGYVLIANOVROYPLQRLRIYRGTQLFEDNALAVLNG 120
Qy      121 DPLNNTPTVGTASPGELRELOLRSLTEILKSGVLIQRNPOLCYODTILMKDIFHNKOLA 180
Db      121 DPLNNTPTVGTASPGELRELOLRSLTEILKSGVLIQRNPOLCYODTILMKDIFHNKOLA 180
Qy      121 LTLIDTNRSRACHPCSPCKSGRCGSESEDCQSLTRVCAAGCARGPLPTDCCHBOC 240
Db      121 LTLIDTNRSRACHPCSPCKSGRCGSESEDCQSLTRVCAAGCARGPLPTDCCHBOC 240
Qy      241 AAGCTGPGHSDCLALCFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Db      241 AAGCTGPGHSDCLALCFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPILHNOEVTADGTORCEKSKPCARVYCYGLMOYIRANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPILHNOEVTADGTORCEKSKPCARVYCYGLMOYIRANSKFIGIT 360
Qy      361 ELERAGCKKIFGSLAFIPESFDGPASNTAALQEPOLQVFTLEITGYLISAMPDLP 420
Db      361 IQEFAAGCKKIFGSLAFIPESFDGPASNTAALQEPOLQVFTLEITGYLISAMPDLP 420
Qy      421 DLSVFQNLQVIRGILHNHGAISLTLOGLIGIWMGLRSRLREGSGGALAHNHTHLCFYHTV 480
Db      421 DLSVFQNLQVIRGILHNHGAISLTLOGLIGIWMGLRSRLREGSGGALAHNHTHLCFYHTV 480
Qy      481 PMDLFRPNFTVSFWLVRPVSAHSHEECVGBGLACHQLCARGHCWGPPTOCVNCQSF 540
Db      481 PMDLFRPNFTVSFWLVRPVSAHSHEECVGBGLACHQLCARGHCWGPPTOCVNCQSF 534
Qy      541 LRGOECVECRVLOGLPREYVNAHCLPCHRECCPONGSVTCRFPADQCYACAHYKDP 600
Db      541 LRGOECVECRVLOGLPREYVNAHCLPCHRECCPONGSVTCRFPADQCYACAHYKDP 594
Qy      601 FCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCHTSCVDLDDKCGPAERASPLTSI 660

```

```
Db      |||||
595 FCVACPSGVKPDLSMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 654
Qy      |||||
661 USAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPTLPGAMPNOQMKI 720
Db      |||||
655 ISAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPTLPGAMPNOQMKI 714
Qy      |||||
721 LKETELRKVKVLGSGAFGTVMKGIWIPDGENYKIPVAIVLRENTSPKANKELIDEAYVM 780
Db      |||||
715 LKETELRKVKVLGSGAFGTVMKGIWIPDGENYKIPVAIVLRENTSPKANKELIDEAYVM 774
Qy      |||||
761 AGVSGPYVSRLLIGICTSTVQLVTOLMPYGCCLLDHVRNRRGLSODLLNMCQIAKGS 840
Db      |||||
775 AGVSGPYVSRLLIGICTSTVQLVTOLMPYGCCLLDHVRNRRGLSODLLNMCQIAKGS 834
Qy      |||||
841 YLEDRLVHRDLAARNVLVKSNNHYKITDPGLARLLDIDETRYHADGKVPKIMWALESI 900
Db      |||||
835 YLEDRLVHRDLAARNVLVKSNNHYKITDPGLARLLDIDETRYHADGKVPKIMWALESI 894
Qy      |||||
901 LRRRFTHOSDVMVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVMI 960
Db      |||||
895 LRRRFTHOSDVMVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVMI 954
Qy      |||||
961 MYKCMIMIDSECRPRRELVSERSMARDPQRFVITQNEIDLGPASPLDSTFFYSLLEDDM 1020
Db      |||||
955 MYKCMIMIDSECRPRRELVSERSMARDPQRFVITQNEIDLGPASPLDSTFFYSLLEDDM 1014
Qy      |||||
1021 GDLVDAEFLVPOQGFPCDPAPAGAGMHHRRSSSTSGGDLTLGLSPSEERPRSP 1080
Db      |||||
1015 GDLVDAEFLVPOQGFPCDPAPAGAGMHHRRSSSTSGGDLTLGLSPSEERPRSP 1074
Qy      |||||
1081 LAPSEGAGSDVFDGDLGMAAGKLOSLPTHDSPLQRYSEDPVPLPSTEDGVAPLTS 1140
Db      |||||
1075 LAPSEGAGSDVFDGDLGMAAGKLOSLPTHDSPLQRYSEDPVPLPSTEDGVAPLTS 1134
Qy      |||||
1141 POPEYVNOVDVAPQPPSPREGPLPAARPAATLERAKTILSPGKNGVKDVFAFGAVENP 1200
Db      |||||
1135 POPEYVNOVDVAPQPPSPREGPLPAARPAATLERAKTILSPGKNGVKDVFAFGAVENP 1194
Qy      |||||
1201 EYLTQGGGAPOPHPPAPAFDNLVYWDODPPERGAPSTPFKGTPTAENPEYLGIDVP 1260
Db      |||||
1195 EYLTQGGGAPOPHPPAPAFDNLVYWDODPPERGAPSTPFKGTPTAENPEYLGIDVP 1254
Qy      |||||
1261 V 1261
Db      |||||
1255 V 1255

RESULT 9
AAB85458
ID      AAB85458 standard; Protein; 1255 AA.
AC      AAB85458;
DT      25-SEP-2001 (first entry)
XX      Human HER-2/neu protein.
XX      Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX      oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
OS      Homo sapiens.
XX      WO200153463-A2.
XX      PD      26-JUL-2001.
XX      PF      19-JAN-2001; 2001WO-US01850.
XX      PR      21-JAN-2000; 2000US-0177545.
XX      PA      (CORI-) CORIXA CORP.
XX
```

```
PI      Cheever MA, Hand-Zimmermann S;
XX      MPI: 2001-476112/51.
DR      N-PSDB; AAH23392.
XX      New antigen-presenting cells, useful as vaccines for eliciting or
PT      enhancing an immune response to HER-2/neu protein, particularly useful
PT      for treating or preventing cancer, e.g. breast cancer -
XX      Claim 2; Page 41-46; 49pp; English.
XX      The invention provides an isolated antigen-presenting cell, which
CC      expresses at least an immunogenic portion of a polypeptide that produces
CC      an immune response to HER-2/neu protein. The antigen-presenting cells are
CC      useful as vaccines for eliciting or enhancing an immune response to
CC      HER-2/neu protein, particularly in treating or preventing malignancies in
CC      which the HER-2/neu oncogene is associated. Specifically, these are
CC      useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC      colon, lung or prostate cancers. The present sequence represents
CC      the human HER-2/neu protein (also known as p185 or c-erbB2).
XX      Sequence 1255 AA;
SQ      Query Match 96.8%; Score 6626; DB 22; Length 1255;
        Best Local Similarity 97.1%; Pred. No. 0;
        Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
Qy      |||||
1 MELALCRWGLLALLPPAASTOVCTGDMKRLPASPTLDMRLHYOGCOVVOGML 60
Db      |||||
1 MELALCRWGLLALLPPAASTOVCTGDMKRLPASPTLDMRLHYOGCOVVOGML 60
Qy      |||||
61 ELTYLPTNANSLFLDIOEVQGVYLIANQVNOVPLQRIYRGTOLEFDNVALAVLDNG 120
Db      |||||
61 ELTYLPTNANSLFLDIOEVQGVYLIANQVNOVPLQRIYRGTOLEFDNVALAVLDNG 120
Qy      |||||
121 DPLNNTPTVGTASPGGLRELQRLSTELIKGVLIQRNPOLCYODTILMKDIFHKNOQLA 180
Db      |||||
121 DPLNNTPTVGTASPGGLRELQRLSTELIKGVLIQRNPOLCYODTILMKDIFHKNOQLA 180
Qy      |||||
181 LTLITNNSRAHPCSPGKSGRCWGESSEDCOSLTRVYACAGCARKCPRLPTDCHEOC 240
Db      |||||
181 LTLITNNSRAHPCSPGKSGRCWGESSEDCOSLTRVYACAGCARKCPRLPTDCHEOC 240
Qy      |||||
241 AAGCTGPKMSDCLACLFHNSGICELHCAALVTYNTDFESMNPNEGRTTFCASCTAPR 300
Db      |||||
241 AAGCTGPKMSDCLACLFHNSGICELHCAALVTYNTDFESMNPNEGRTTFCASCTAPR 300
Qy      |||||
301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db      |||||
301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Qy      |||||
361 ELIEFAGCKKIFGSLAFLESFPGDPASNTAPLQPELOLVFETLEITGYLIYISAMPDILP 420
Db      |||||
361 IDEFAGCKKIFGSLAFLESFPGDPASNTAPLQPELOLVFETLEITGYLIYISAMPDILP 420
Qy      |||||
421 DLSVFQNTQVTRGRLIHNGAVSLTLQGLISWLGRLSRLBEGSSGALILHNNHLCVPHV 480
Db      |||||
421 DLSVFQNTQVTRGRLIHNGAVSLTLQGLISWLGRLSRLBEGSSGALILHNNHLCVPHV 480
Qy      |||||
481 PWDQLFRNNFTVSFWLVPKVASHLBECVGEGLAQHLCARGHCWGGPTQCVNCSQF 540
Db      |||||
481 PWDQLFRNNFTVSFWLVPKVASHLBECVGEGLAQHLCARGHCWGGPTQCVNCSQF 534
Qy      |||||
541 LRQDECECRVLOQLPREYVNAHRLPCHPECOQONGSVTCFGEADQCVACAHYKDBP 600
Db      |||||
535 LRQDECECRVLOQLPREYVNAHRLPCHPECOQONGSVTCFGEADQCVACAHYKDBP 594
Qy      |||||
601 FCVACPSGVKPDLSMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 660
Db      |||||
595 FCVACPSGVKPDLSMPIMKPPDEGACQPCPINCTHSCVDLDKGCBAEORASPLTSI 654
Qy      |||||
661 USAVVGILLVVLGVVFGILIKRROOKIRKYMRLLOETELVEPTLPGAMPNOQMKI 720
```

```

Db      655 ISAVVGLLVVGVFGLIKRRQOKIRKYMRLQETELVEPLTPSGAMPNOAQMRI 714
Qy      721 LKETELRKVKVLSGAGFGTVKGIIMPDGENVKIPVAIKVLRNTSPKANEIDEAVYM 780
      715 LKETELRKVKVLSGAGFGTVKGIIMPDGENVKIPVAIKVLRNTSPKANEIDEAVYM 774
Qy      781 AGVGSPPVYSRLGLICTSTVQVLTQMPYGCGLDHVRENRGLSQDILNMCQIAKMS 840
      775 AGVGSPPVYSRLGLICTSTVQVLTQMPYGCGLDHVRENRGLSQDILNMCQIAKMS 834
Qy      841 YLEEVRLVHRDLAARNLVKSPNHNKITTDEGLARLLDIDETRYHADGKVPKIMMALESI 900
      835 YLEEVRLVHRDLAARNLVKSPNHNKITTDEGLARLLDIDETRYHADGKVPKIMMALESI 894
Qy      901 LRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPICTIDYMI 960
      895 LRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPICTIDYMI 954
Qy      961 MVKCMIDSECRPRFRELVSFSSMARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDMD 1020
      955 MVKCMIDSECRPRFRELVSFSSMARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDMD 1014
Qy      1021 GDLVDAEYLVPOGFFCPCDPAPAGAGVNHRRSSSTRSGGDLTLGLPSEBEAPRSP 1080
      1015 GDLVDAEYLVPOGFFCPCDPAPAGAGVNHRRSSSTRSGGDLTLGLPSEBEAPRSP 1074
Qy      1081 LAPSEGASDVFDGDLGKAAGKGLQSLPTHDPSPLORYSEDPVPLPSETGYVAPLTC 1140
      1075 LAPSEGASDVFDGDLGKAAGKGLQSLPTHDPSPLORYSEDPVPLPSETGYVAPLTC 1134
Qy      1141 POPEYVNPQVPRPQPSPREGPPLPAPAGATLERAKTLSPGKGVNVDAVFAFGAVENP 1200
      1135 POPEYVNPQVPRPQPSPREGPPLPAPAGATLERAKTLSPGKGVNVDAVFAFGAVENP 1194
Qy      1201 EYLTPOGGAAPQPPPPAFSAFNLTYMDODPPERGAPESTFGTPTAENPEYLGIDVP 1260
      1195 EYLTPOGGAAPQPPPPAFSAFNLTYMDODPPERGAPESTFGTPTAENPEYLGIDVP 1254
Qy      1261 V 1261
      1255 V 1255
Db

```

```

RESULT 10
AAC88267
ID AAG88267 standard; Protein; 1255 AA.
AC AAG88267;
DT 11-SEP-2001 (first entry)
DE HER2/neu amino acid sequence.
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
  immune response; vaccine; tumour; cancer; cytotoxic T lymphocyte; CTL.
  tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
PN WO200141787-A1.
PD 14-JUN-2001.
PF 11-DEC-2000; 2000WO-US33591.
PR 10-DEC-1999; 99US-0458299.
XX (EPIM-) EPIMMUNE INC.
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX PI Keogh E;
XX MPI; 2001-374995/39.

```

```

XX      An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT      cellular immune responses for the prevention and treatment of cancer -
XX
PS      Disclosure; Page 15; 1999; English.
XX
CC      The present invention describes isolated prepared HER2/neu epitopes (I).
CC      Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC      culture in vitro and binds to a complex of an epitope (I), bound to a
CC      human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC      and a second epitope and the peptide is less than 50 contiguous amino
CC      acids that have 100% identity with a native peptide sequence of HER2/neu;
CC      (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC      excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC      (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC      and immunostimulant activities, and can be used in vaccines. (I), (II)
CC      and (III) are useful for inducing cellular immune responses for the
CC      prevention and treatment of cancer. (I) and (II) are useful for
CC      monitoring or evaluating an immune response to a tumour-associated
CC      antigen when incubated with a T lymphocyte sample from a patient and
CC      detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC      based vaccines mean that immunosuppressive epitopes that may be present
CC      in whole antigens may be avoided. Selected epitopes may be combined to
CC      enhance immunogenicity. The possible pathological side effects caused by
CC      infectious agents or whole protein antigen is eliminated. The vaccine
CC      provides the ability to direct and focus an immune response to multiple
CC      selected antigens from the same pathogen. Epitope-based anti-tumour
CC      vaccines provides the opportunity to combine epitopes derived from
CC      multiple tumour-associated molecules addressing the problem of tumour-
CC      tumour variability and reducing the likelihood of tumour escape due to
CC      antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC      the exemplification of the present invention.
SQ      Sequence 1255 AA;
Query Match 96.8%; Score 6626; DB 22; Length 1255;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
Qy      1 MELAALCRWGLLLALPPGAASVQCTGDMKRLPASPEHLDMLRHLVGGCQVQGNL 60
      1 MELAALCRWGLLLALPPGAASVQCTGDMKRLPASPEHLDMLRHLVGGCQVQGNL 60
Db
Qy      61 ELTYLPTNASLSPLODIOEVQGYVLIANOVQVPLQRLIVRGOLFEDNYALAVDNG 120
      61 ELTYLPTNASLSPLODIOEVQGYVLIANOVQVPLQRLIVRGOLFEDNYALAVDNG 120
Db
Qy      121 DPLNNTTPVTGASPGRLRELQRLSTELIKGVLIQRNPOLCYOPTILMKDIFHNQOLA 180
      121 DPLNNTTPVTGASPGRLRELQRLSTELIKGVLIQRNPOLCYOPTILMKDIFHNQOLA 180
Db
Qy      181 LTLIDNRSRACHPCSPWCKSRGSGSSSDQSLTRTVAGGACRCKGPPPTDCHEQC 240
      181 LTLIDNRSRACHPCSPWCKSRGSGSSSDQSLTRTVAGGACRCKGPPPTDCHEQC 240
Db
Qy      241 AAGCTGKHSQCLACHFNHSGICELCPALVYNTDTFESMPNPEGYTTGASCVTACP 300
      241 AAGCTGKHSQCLACHFNHSGICELCPALVYNTDTFESMPNPEGYTTGASCVTACP 300
Db
Qy      301 YNYLSTVGSCTVLCPLHNOEVTAEQDGTORCKSKPCARVCYGLGMQYIRANSKFIGIT 360
      301 YNYLSTVGSCTVLCPLHNOEVTAEQDGTORCKSKPCARVCYGLGMQYIRANSKFIGIT 360
Db
Qy      361 ELEFAGCKKIIGSLAFLPESFDGDPASNTAFLQEOLOVFTLEITGYLYISAMPDLP 420
      361 IOEFAGCKKIIGSLAFLPESFDGDPASNTAFLQEOLOVFTLEITGYLYISAMPDLP 420
Db
Qy      421 DLSVFOVLQVIRGILLNGAYSLTLOGISWLGRLSLRELGSGLAIHNRTHLCFVHTV 480
      421 DLSVFOVLQVIRGILLNGAYSLTLOGISWLGRLSLRELGSGLAIHNRTHLCFVHTV 480
Db
Qy      481 PMDQLEFRNFTVSWFLRVKVSASHLEECVGEGLACHQLCARGHCGPGTQCVCNSQF 540
      481 PMDQLEFRNFTVSWFLRVKVSASHLEECVGEGLACHQLCARGHCGPGTQCVCNSQF 540

```

```

Db      481  PMDOLFRNPHOALLHTANRP-----DECVBGLACHQLCARGHCWBGPTQCVNCSQF 534
Qy      541  LRGOECVEBCRVLOGLPREYVNAARHCLPCHPECQPNQSVTCFGEADOCVACAHYKDDP 600
Db      535  LRGOECVEBCRVLOGLPREYVNAARHCLPCHPECQPNQSVTCFGEADOCVACAHYKDDP 594
Qy      601  FCVACPCSGVKDPLSYMPIWKFPPDEGACQPCPINCNSCVDLDKGCBAEORASPLTSI 660
Db      595  FCVACPCSGVKDPLSYMPIWKFPPDEGACQPCPINCNSCVDLDKGCBAEORASPLTSI 654
Qy      661  VSAVVGILLVVVLGVFGLIKRROOKIKKTYMRLLQSTELVEPLTBSGANPNQAKM1 720
Db      655  ISAVVGILLVVVLGVFGLIKRROOKIKKTYMRLLQSTELVEPLTBSGANPNQAKM1 714
Qy      721  LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKYLBRENTSPKANKELIDAYYM 780
Db      715  LKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKYLBRENTSPKANKELIDAYYM 774
Qy      781  AGVSPYVSRLLIGICTSTVOLVTOLMPYGCLLDHVRENRGRIGSODLLNMCQIAKMS 840
Db      775  AGVSPYVSRLLIGICTSTVOLVTOLMPYGCLLDHVRENRGRIGSODLLNMCQIAKMS 834
Qy      841  YLEDVRLVHRDLAARNVLVKSPPNHVKITDGLARLLDIDETVHADGGKVP1KMMALBS1 900
Db      835  YLEDVRLVHRDLAARNVLVKSPPNHVKITDGLARLLDIDETVHADGGKVP1KMMALBS1 894
Qy      901  LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMI 960
Db      895  LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMI 954
Qy      961  MYKCMIDSECRPRRELVSERFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM 1020
Db      955  MYKCMIDSECRPRRELVSERFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM 1014
Qy      1021  GDVLVAEEVLVQOGFPCDPAPGAGGMVHHRSSSTSGGDLTLGISEPEEERPRSP 1080
Db      1015  GDVLVAEEVLVQOGFPCDPAPGAGGMVHHRSSSTSGGDLTLGISEPEEERPRSP 1074
Qy      1081  LAPSEGASDVEGDGLGMAKAGLOSLPTHDSPLQRYSEDPVPLPSESTDGVAPLTS 1140
Db      1075  LAPSEGASDVEGDGLGMAKAGLOSLPTHDSPLQRYSEDPVPLPSESTDGVAPLTS 1134
Qy      1141  POPEVYNQDVARPOPPSPREGPLPAARPGATLERAKTSLSPKNGVVKDVFAGAVENP 1200
Db      1135  POPEVYNQDVARPOPPSPREGPLPAARPGATLERAKTSLSPKNGVVKDVFAGAVENP 1194
Qy      1201  EYLTFQGAAPQPHPPAPAFNDLXYWDODPPERGAPSTFFKGTPTAENBEYLGDDVP 1260
Db      1195  EYLTFQGAAPQPHPPAPAFNDLXYWDODPPERGAPSTFFKGTPTAENBEYLGDDVP 1254
Qy      1261  V 1261
Db      1255  V 1255

```

RESULT 11

AAE24067 ID AAE24067 standard; Protein; 1255 AA.

AAE24067; AC

23-SEP-2002 (first entry)

Human Her-2 protein.

Human: Her-2; epidermal growth factor receptor 2; infection; cancer;

hyperproliferative disorder; prophylaxis; inflammation; antisense;

tumour; gene therapy; phosphothioate backbone.

Homc. sapiens.

MO200222636-A1.

```

PD      21-MAR-2002.
XX
XX      12-SEP-2001; 2001WO-US28572.
PF
PR      15-SEP-2000; 2000US-0663834.
XX
XX      (ISIS-) ISIS PHARM INC.
PA
PI      Bennett CF, Cowse LM;
XX
XX      WPI; 2002-471192/50.
DR      N-PSDB; AAD38904.
XX
XX      Novel antisense oligonucleotide which modulates the expression of Human
PT      Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT      inflammation or to prevent infection in humans.
XX
XX      Example 13; Page 95-107; 116pp; English.
XX
XX      The invention relates to antisense compounds targeted to a nucleic
XX      acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX      that specifically hybridises with and inhibits the expression of Her2.
XX      Antisense compounds of the invention are used for treating diseases or
XX      conditions associated with Her2 such as hyperproliferative disorders
XX      e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX      neural or cardiac cancer. They are also useful prophylactically e.g.
XX      to prevent or delay infection, inflammation and tumour formation. The
XX      invention is also used in gene therapy. The present sequence is human
XX      Her-2 protein.

```

Sequence 1255 AA;

Query Match 96.8%; Score 6626; DB 23; Length 1255;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;

```

Qy      1  MELALACRWGLLALLPQAASTQVCTGDMRLRPASPTHLDMLRHLVYQCCQVYQGNL 60
Db      1  MELALACRWGLLALLPQAASTQVCTGDMRLRPASPTHLDMLRHLVYQCCQVYQGNL 60
Qy      61  ELTYLPTNASLSFLDIOEVQGYVLIANNQVQVPLQRLIRVGTQLEFEDNYALAVLDNG 120
Db      61  ELTYLPTNASLSFLDIOEVQGYVLIANNQVQVPLQRLIRVGTQLEFEDNYALAVLDNG 120
Qy      121  DPLNNTTPVTGASPGGLAELOLRSLTEILKGVLIQRNPQLCYODTILMKDIFHRNQOLA 180
Db      121  DPLNNTTPVTGASPGGLAELOLRSLTEILKGVLIQRNPQLCYODTILMKDIFHRNQOLA 180
Qy      181  LTLIDTNSRACHPCSPMKCKSGRCSGESSEDCQSLTRTYCAGACGACRCKGPLPTDCHEQC 240
Db      181  LTLIDTNSRACHPCSPMKCKSGRCSGESSEDCQSLTRTYCAGACGACRCKGPLPTDCHEQC 240
Qy      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300
Db      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300
Qy      301  YNYLSTDVGSCTLVCPILNQEVTAEDGTQRCCKSKPCARVCYGIAMQYIKANSKFIGIT 360
Db      301  YNYLSTDVGSCTLVCPILNQEVTAEDGTQRCCKSKPCARVCYGIAMQYIKANSKFIGIT 360
Qy      361  ELEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQAPQLOVFEETLEETIGLYLSAMPDSL 420
Db      361  IOFAGCKKIFGSLAFLEPESFDGDPASNTAPLQAPQLOVFEETLEETIGLYLSAMPDSL 420
Qy      421  DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGARSRELGSGALLIHNHTLFCFVHTV 480
Db      421  DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGARSRELGSGALLIHNHTLFCFVHTV 480
Qy      481  PMDOLFRNPNFTVSWLVRPVKSASHLEECVGBGLACHQLCARGHCWBGPTQCVNCSQF 540
Db      481  PMDOLFRNPHOALLHTANRP-----DECVBGLACHQLCARGHCWBGPTQCVNCSQF 534
Qy      541  LRGOECVEBCRVLOGLPREYVNAARHCLPCHPECQPNQSVTCFGEADOCVACAHYKDDP 600

```

```

Db      535  LRGECEVECRVLCGLPREYVNAHCHLPCHPEQONSGVTCFGBADQVCAHAKDDP 594
Qy      601  FCVARCPGKVPKDISYMPIMKFPDEGACQPCPINCSTCVLDKDGCPAEORASPLTSI 660
Db      595  FCVARCPGKVPKDISYMPIMKFPDEGACQPCPINCSTCVLDKDGCPAEORASPLTSI 654
Qy      661  VSAVVGILLVVLGVVFGILLKRRQOKIRKYMRLQETELVEPLTSGGAMPNOQMRI 720
Db      655  ISAVVGILLVVLGVVFGILLKRRQOKIRKYMRLQETELVEPLTSGGAMPNOQMRI 714
Qy      721  LKETELRKVKYLGSAGFTVYKGIWIPQGENYKIVAIKVLRENTSPKXNKELIEAYVM 780
Db      715  LKETELRKVKYLGSAGFTVYKGIWIPQGENYKIVAIKVLRENTSPKXNKELIEAYVM 774
Qy      781  AGVGSPIYVSRLLGICLTSTVQLVTLQMPYGLLDHVRNKGRLSGODLLNMCQIAKMS 840
Db      775  AGVGSPIYVSRLLGICLTSTVQLVTLQMPYGLLDHVRNKGRLSGODLLNMCQIAKMS 834
Qy      841  YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETBYHADGKVPKMMALBSI 900
Db      835  YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETBYHADGKVPKMMALBSI 894
Qy      901  LRRPFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRLEPPPICTIDVMI 960
Db      895  LRRPFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRLEPPPICTIDVMI 954
Qy      961  MYKCMIDSECRPRERELVSFESRWARDPQRFVVIQNEDLGPASPLDSTFYRSLDEDDM 1020
Db      955  MYKCMIDSECRPRERELVSFESRWARDPQRFVVIQNEDLGPASPLDSTFYRSLDEDDM 1014
Qy      1021  GDLYDAEYLVPOQCFPCPDPAKPGAGVNHHRSSSTRSGGDLTLGLEPSEEBAPRSP 1080
Db      1015  GDLYDAEYLVPOQCFPCPDPAKPGAGVNHHRSSSTRSGGDLTLGLEPSEEBAPRSP 1074
Qy      1081  LAPSEGASDVDFDGLGKAAKGLQSLPTHDPSPLOXYSEDPVPLPSETOGVYVPLTCS 1140
Db      1075  LAPSEGASDVDFDGLGKAAKGLQSLPTHDPSPLOXYSEDPVPLPSETOGVYVPLTCS 1134
Qy      1141  POPEYVNPQVVRPOPSPREGRPLPAARPAATLEBRAKTLSPKNGVNDVFAFGAVENP 1200
Db      1135  POPEYVNPQVVRPOPSPREGRPLPAARPAATLEBRAKTLSPKNGVNDVFAFGAVENP 1194
Qy      1201  EYLTPQGGAAQPPHPPAFSAFNDLYWDDPPERGAPESTFKCTPTAENPEYIGLDVP 1260
Db      1195  EYLTPQGGAAQPPHPPAFSAFNDLYWDDPPERGAPESTFKCTPTAENPEYIGLDVP 1254
Qy      1261  V 1261
Db      1255  V 1255

RESULT 12
AAE20479
ID      AAE20479 standard; Protein: 1255 AA.
XX
AC      AAE20479;
XX
DT      01-JUL-2002 (first entry)
XX
DE      Human Her-2/neu protein.
XX
KW      Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX      human leukocyte antigen; HLA; vaccine; malignancy; cytosolic.
XX      Homo sapiens.
XX
FH      Key
FT      Region
XX      Location/Qualifiers
XX      1021..1030
XX      /note= "Naturally processed HLA-B*44-restricted epitope"
XX      WO200214503-A2.

```

```

PD      21-FEB-2002.
XX
XX      14-AUG-2001; 2001WO-US41733.
XX
XX      14-AUG-2000; 2000US-225152P.
XX      28-SEP-2000; 2000US-236428P.
XX      21-FEB-2001; 2001US-270520P.
XX
XX      (CORI-) CORIXA CORP.
XX
XX      Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD,
XX      McNeill PD, Vedralick TS;
XX      WPI; 2002-280758/32.
XX      N-PSDB; AAD32743.
XX
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX      prevention and diagnosis of cancer, preferably breast cancer
XX
XX      Disclosure; Page 114-117; 129pp; English.
XX
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
XX      eliciting an immune response in a patient, where the patient is human
XX      leukocyte antigen (HLA)-B*44 positive or is affected with breast cancer.
XX      The composition is useful for the therapy and diagnosis of cancer,
XX      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX      and other compositions for the diagnosis, prevention and treatment of
XX      human malignancies, for stimulating and/or expanding T cells specific for
XX      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX      patient. The invention is useful for stimulating a T cell response in a
XX      human patient, as probe or primer for nucleic acid hybridisation, to
XX      selectively form duplex molecules with complementary stretches of the
XX      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX      length gene from a suitable library, and to direct expression of a
XX      polypeptide in appropriate host cells. The composition is useful in
XX      prophylactic or therapeutic applications and for the treatment of cancer,
XX      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
XX
XX      Sequence 1255 AA:
XX
XX      Query Match      96.8%; Score 6626; DB 23; Length 1255;
XX      Best Local Similarity 97.1%; Pred. No. 0;
XX      Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
XX
Qy      1  MELAALCRWGLLALIPGAASVQCTGTDMLRLPASPTHLDMLRLHYQCCQVQGNL 60
Db      1  MELAALCRWGLLALIPGAASVQCTGTDMLRLPASPTHLDMLRLHYQCCQVQGNL 60
Qy      61  ELYLPTNASSPLQDIOEVQGVYLIANOVROVPLQRLRYRGQLFEDNVALAVLDNG 120
Db      61  ELYLPTNASSPLQDIOEVQGVYLIANOVROVPLQRLRYRGQLFEDNVALAVLDNG 120
Qy      121  DPLNNTTPVTGASPGSLRELOLRSLTEILKGGVLIQRNPOLCYOQDTIMKIDFHKNOQLA 180
Db      121  DPLNNTTPVTGASPGSLRELOLRSLTEILKGGVLIQRNPOLCYOQDTIMKIDFHKNOQLA 180
Qy      181  LTLIDTNRSRACHPCSPKCKSRCKGSESSDQSLTRTVCAAGCARCKGRLPTDCCHQC 240
Db      181  LTLIDTNRSRACHPCSPKCKSRCKGSESSDQSLTRTVCAAGCARCKGRLPTDCCHQC 240
Qy      241  AAGCTGPRGSDCLAHNHSIGIELHCPALVTVNTDFFESMPNPEGYTTGASCVTACP 300
Db      241  AAGCTGPRGSDCLAHNHSIGIELHCPALVTVNTDFFESMPNPEGYTTGASCVTACP 300
Qy      301  YNVLSTVGSCCLVCPHLNDEVTADGTORCEKSKPCARVCYGLAMQYIYANSKFIGIT 360
Db      301  YNVLSTVGSCCLVCPHLNDEVTADGTORCEKSKPCARVCYGLAMQYIYANSKFIGIT 360
Qy      361  ELEFAGCKKIFGSLAFLPESFDGPASNTAPLQPEQLQVETLEITGYLYISAMPDSL 420
Db      361  ELEFAGCKKIFGSLAFLPESFDGPASNTAPLQPEQLQVETLEITGYLYISAMPDSL 420

```

Db	361	IOEFAAGCKKI	FOSLAFLFEPSPDGDASNPAPLQPELDQVFEETLLEETGLVYISANPDISLP	420
QY	421	DLVSFQNLQVYIGRIILHNGAYSLLTQGLGISWLGRLSRELGSGALIHNNHLCEVHTV		480
Db	421	DLVSFQNLQVYIGRIILHNGAYSLLTQGLGISWLGRLSRELGSGALIHNNHLCEVHTV		480
QY	481	PMDOLEFRFNFPVVSFWLRLVPKVSASHLSECEVGEGLACHOLCARGHOMGGPQCNVCSQF		540
Db	481	PMDOLEFRFNFPVVSFWLRLVPKVSASHLSECEVGEGLACHOLCARGHOMGGPQCNVCSQF		540
QY	541	LARGOCEVECEVYLOGLPREYVNAARHCLPCHPECPQPNQSVTCFGEADQCVACAHYKDP		600
Db	535	LARGOCEVECEVYLOGLPREYVNAARHCLPCHPECPQPNQSVTCFGEADQCVACAHYKDP		594
QY	601	FCVNAACPSGVKNDLSYMPIMKFPDEGACQPCPINCTHSCVLDLDDKCPAEBRASPILSI		660
Db	595	FCVNAACPSGVKNDLSYMPIMKFPDEGACQPCPINCTHSCVLDLDDKCPAEBRASPILSI		654
QY	661	VSAAVVGILLVVLGVVFGILLIKRROOKIKRYTMRRLLQSTELVEPLTPSGAMPNOAMRI		720
Db	655	ISAVVGILLVVLGVVFGILLIKRROOKIKRYTMRRLLQSTELVEPLTPSGAMPNOAMRI		714
QY	721	LKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDEAYVM		780
Db	715	LKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDEAYVM		774
QY	781	AGVGSPPVYSRLIGLICLTSTVOLVTOLMRPGCCLDHYRENRGLSGODLLNMCQIAKGS		840
Db	775	AGVGSPPVYSRLIGLICLTSTVOLVTOLMRPGCCLDHYRENRGLSGODLLNMCQIAKGS		834
QY	841	YLEDEVRLVHRDLAARNVLVKSBNHVKITDPFGLARLIDIDETRYHADGKVPKIMMALESI		900
Db	835	YLEDEVRLVHRDLAARNVLVKSBNHVKITDPFGLARLIDIDETRYHADGKVPKIMMALESI		894
QY	901	LRRRTTHQSDWVSYGVTWELMTFGAKPRPDGIPAEIIPDLEKGERLPOPPICITIDVYMI		960
Db	895	LRRRTTHQSDWVSYGVTWELMTFGAKPRPDGIPAEIIPDLEKGERLPOPPICITIDVYMI		954
QY	961	MVKCMIMDISCECPRRRELVSERFSRMAARDPQRFVVIQNEBLSGPASPLDSTFFYSLLIEDDM		1020
Db	955	MVKCMIMDISCECPRRRELVSERFSRMAARDPQRFVVIQNEBLSGPASPLDSTFFYSLLIEDDM		1014
QY	1021	GDVLDAEELVLPQGFPCPDPAFGAGVNHHRSSSTRSGGDLTLGLEPSEBEAPRSP		1080
Db	1015	GDVLDAEELVLPQGFPCPDPAFGAGVNHHRSSSTRSGGDLTLGLEPSEBEAPRSP		1074
QY	1081	LAPSRGASDVDEGLGMAAGLOSLPTHDSPLQKRSSEDPYVLPBETDGYVAPLTC		1140
Db	1075	LAPSRGASDVDEGLGMAAGLOSLPTHDSPLQKRSSEDPYVLPBETDGYVAPLTC		1134
QY	1141	POPEYVNPQDVRPQPPSPREGRLPAARAGATLEBAKTLSPGKNGVNVDOVFAGGAVENP		1200
Db	1135	POPEYVNPQDVRPQPPSPREGRLPAARAGATLEBAKTLSPGKNGVNVDOVFAGGAVENP		1194
QY	1201	EYLTTPQGGAAQPPHPPPAPSPAFNDLYYWDODPPERGAAPSTFKGTPTAENPEYLGIDVP		1260
Db	1195	EYLTTPQGGAAQPPHPPPAPSPAFNDLYYWDODPPERGAAPSTFKGTPTAENPEYLGIDVP		1254
QY	1261	V 1261		
Db	1255	V 1255		
RESULT 13				
AAMS1143				
ID	AAMS1143	standard; Protein; 1255 AA.		
XX	AAMS1143;			
AC				
XX	17-JUN-2002	(first entry)		
XX	Human Her-2/neu oncogene-encoded p185 glycoprotein.			

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 XM tyrosine kinase; receptor; c-erbB2; gene therapy.
 OS Homo sapiens.
 PT Key Location/Qualifiers
 FH Domain 1..653
 FT /note= "extracellular domain"
 FT 676..1255
 FT /note= "intracellular domain"
 FT Domain 990..1255
 FT /note= "phosphorylation domain"
 PN MO200212341-A2.
 XX 14-FEB-2002.
 PD 03-AUG-2001; 2001WO-US24283.
 PF 03-AUG-2001; 2000US-0632507.
 PR 03-AUG-2000; 2000US-0632507.
 PA (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Cheever MA, Gheysen D;
 PI MPI. 2002-241743/29.
 DR N-PSDB; ABA92250.
 DR Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain -
 XX
 PS Claim 68; Fig 7; 141pp; English.
 XX
 CC The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines
 CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its DeltaCD fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC administering cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.
 XX
 SQ Sequence 1255 AA;
 Query Match 96.8%; Score 6626; DB 23; Length 1255;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1.
 1 MELAAALRWGILLALLPRGAASVGVCTGDKRLPLASPEHMLRLYLGGCGVAGNL 60
 1 MELDAALRWGILLALLPRGAASVGVCTGDKRLPLASPEHMLRLYLGGCGVAGNL 60


```

QY 61 ELTFLPTNASTSLFLODIOEVGYVLIANNOVQVPLQRLRIYRGTOLEEDNVALAVLNG 120
DB 61 ELTFLPTNASTSLFLODIOEVGYVLIANNOVQVPLQRLRIYRGTOLEEDNVALAVLNG 120
QY 121 DPLNNTTQVGTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTIIMKDFHKNOLA 180
DB 121 DPLNNTTQVGTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYODTIIMKDFHKNOLA 180
QY 181 LTLIDITNRSRACHPSCPMCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLETDCCHQC 240
DB 181 LTLIDITNRSRACHPSCPMCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLETDCCHQC 240
QY 241 AAGCTGPGHSDCLACLHFNHSGICELHCPALVTYNTDTPESPNPNEGRTTFCASCVTACP 300
DB 241 AAGCTGPGHSDCLACLHFNHSGICELHCPALVTYNTDTPESPNPNEGRTTFCASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPFLHNOEVTADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPFLHNOEVTADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 ELLEFAGCKKIFGSLAFLEBSFDGDPASNTAPLQPBOLQVFETLEITGYLIASMPDLP 420
DB 361 IOEFAGCKKIFGSLAFLEBSFDGDPASNTAPLQPBOLQVFETLEITGYLIASMPDLP 420
QY 421 DLSVFQNIQVIRGRILHNGAYSILTQIGISWLGRLSRLGSGIALIHNTHLCFVHTV 480
DB 421 DLSVFQNIQVIRGRILHNGAYSILTQIGISWLGRLSRLGSGIALIHNTHLCFVHTV 480
QY 481 PMDOLFRFNFTVSWLTVPKVSASHLEECVGEGLACHQOLCARHCWMPGPGPCNCSOF 540
DB 481 PMDOLFRFNFTVSWLTVPKVSASHLEECVGEGLACHQOLCARHCWMPGPGPCNCSOF 540
QY 541 LRGOECVEECRVLOGLPREYVNAHRLCPHECOPONGSVTCFGEADQCVACAHYKDP 600
DB 541 LRGOECVEECRVLOGLPREYVNAHRLCPHECOPONGSVTCFGEADQCVACAHYKDP 600
QY 601 FCVARCPBGVPRDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDKCGPAEGRASPLTSI 660
DB 601 FCVARCPBGVPRDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDKCGPAEGRASPLTSI 660
QY 661 VSAVVGLILVVVLGVVFGILIKRROQIRKTYMRLLQETELVEPLTSGAMPNOAORI 720
DB 661 VSAVVGLILVVVLGVVFGILIKRROQIRKTYMRLLQETELVEPLTSGAMPNOAORI 720
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKI PVAIKVLRENTSPPRANKEILDEAVM 780
DB 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKI PVAIKVLRENTSPPRANKEILDEAVM 780
QY 781 AGVSPVYSRLGICLTSTVOLVQLMYGGCLLDHVRNRRGLSGODLLNMCQIAKMS 840
DB 781 AGVSPVYSRLGICLTSTVOLVQLMYGGCLLDHVRNRRGLSGODLLNMCQIAKMS 840
QY 841 YLEEDVRLVHRDLAARNVVKSPNHVKITDPGLARLDDI DETEYHADGGKIVIKXMALESI 900
DB 841 YLEEDVRLVHRDLAARNVVKSPNHVKITDPGLARLDDI DETEYHADGGKIVIKXMALESI 900
QY 901 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRPLPQPICTIDVMI 960
DB 901 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGRPLPQPICTIDVMI 960
QY 961 MVKCMWIMDSERPRFRRELVSFSRMAADPORFVYI ONEDLCPASPLDSTFRSLIEDDM 1020
DB 961 MVKCMWIMDSERPRFRRELVSFSRMAADPORFVYI ONEDLCPASPLDSTFRSLIEDDM 1020
QY 1021 GDVLDAEEYLVPOQGFPCPDPAFAGAGVHHRHSSSTRSGGDLTGLLEPSEEBAPSP 1080
DB 1021 GDVLDAEEYLVPOQGFPCPDPAFAGAGVHHRHSSSTRSGGDLTGLLEPSEEBAPSP 1080
QY 1081 LAPSEGASDVDFDGLGKAAGLQSLPTHDPSPLORYSEDPVPLPSETIDGYAAPLTCS 1140
DB 1081 LAPSEGASDVDFDGLGKAAGLQSLPTHDPSPLORYSEDPVPLPSETIDGYAAPLTCS 1140
QY 1141 POPEYVNOPOVVRPPSPREBPLPAARPAATLERAKTSLPGKXGVVXKDVAFGGAIVENP 1200
DB 1141 POPEYVNOPOVVRPPSPREBPLPAARPAATLERAKTSLPGKXGVVXKDVAFGGAIVENP 1200

```

```

DB 1135 POPEYVNOPOVVRPPSPREBPLPAARPAATLERAKTSLPGKXGVVXKDVAFGGAIVENP 1194
QY 1201 EYLTPOGGAAPQHPHPPAFSPAFDNLYYWDDPPRGAAPPSTFKGTPRANEPEYGLDVP 1260
DB 1195 EYLTPOGGAAPQHPHPPAFSPAFDNLYYWDDPPRGAAPPSTFKGTPRANEPEYGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
XX AC AAU77114;
XX AC
XX 05-JUN-2002 (first entry)
XX
XX Human Her-2/neu polypeptide.
XX
XX Human, Her-2/neu; cytostatic; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX
XX Homo sapiens.
XX
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX DR N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
XX encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide -
XX
XX Disclosure; Page 71-74; 74pp; English.
XX
XX The invention relates to a method for inhibiting development of
XX haematological malignancy in a patient by administering a polypeptide
XX comprising an immunogenic portion of Her-2/neu or a polynucleotide
XX encoding the polypeptide. Antigen presenting cells that express the
XX protein can also be administered. The sequences are used for inhibiting
XX development of haematological malignancy such as acute myelogenous
XX leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
XX leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
XX lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
XX Sequence 1255 AA;
XX
XX Query Match 96.8%; Score 6626; DB 23; Length 1255;
XX Best Local Similarity 97.1%; Pred. No. 0;
XX Matches 1224; Conservative 9; Mismatches 22; Indels 6; Gaps 1;
XX
QY 1 METALALRWKGLLALLPRGAASSTOVCTGTWKRLRSPETHLMRLHYGCGVQGNL 60
DB 1 METALALRWKGLLALLPRGAASSTOVCTGTWKRLRSPETHLMRLHYGCGVQGNL 60
QY 61 ELTFLPTNASTSLFLODIOEVGYVLIANNOVQVPLQRLRIYRGTOLEEDNVALAVLNG 120
DB 61 ELTFLPTNASTSLFLODIOEVGYVLIANNOVQVPLQRLRIYRGTOLEEDNVALAVLNG 120

```

```

Db      61  ELTYLPTNASLSFLDIOIEVQGVYLIAHNOVRQVPLQRLRIVRGTQLEFDNYALAVLNG 120
Qy      121  DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNOLA 180
Db      121  DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNOLA 180
Qy      181  LTLIDTNRSRACHPCSPMKCKSRGWGSESSDCCSLTRTVCAAGCARCKGPLPTDCHEQC 240
Db      181  LTLIDTNRSRACHPCSPMKCKSRGWGSESSDCCSLTRTVCAAGCARCKGPLPTDCHEQC 240
Qy      241  AAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDFESNPNEGRTYFGASCVTACP 300
Db      241  AAGCTGPRHSDCLACLHFNHSGICELHCPALVTYNTDFESNPNEGRTYFGASCVTACP 300
Qy      301  YNYLSTDVGSCTLVCPRLHNOEVTABEDGTORCEKSGPCARVCGYGMQVITKANSKFIGIT 360
Db      301  YNYLSTDVGSCTLVCPRLHNOEVTABEDGTORCEKSGPCARVCGYGMQVITKANSKFIGIT 360
Qy      361  ELEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPEOLQVFETLEETGYLYISAMPDSLIP 420
Db      361  IOEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPEOLQVFETLEETGYLYISAMPDSLIP 420
Qy      421  DLSVTONIQTIRGRILLNNGAYSILTQGLGISWLGLRSIRELSGLALIHNTHLCPVHTV 480
Db      421  DLSVTONIQTIRGRILLNNGAYSILTQGLGISWLGLRSIRELSGLALIHNTHLCPVHTV 480
Qy      481  PMDOLFRFNFTVSMLRVPKVSASHLECEVGEGLAQHQLCARGCMGSPQCVCNCSQF 540
Db      481  PMDOLFRFNFTVSMLRVPKVSASHLECEVGEGLAQHQLCARGCMGSPQCVCNCSQF 540
Qy      541  LRGOCEVEECRLQGLPREYVNARHCLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600
Db      541  LRGOCEVEECRLQGLPREYVNARHCLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600
Qy      594  LRGOCEVEECRLQGLPREYVNARHCLPCHPECOPONGSVTCFGEADQCVACAHYKDP 594
Db      594  LRGOCEVEECRLQGLPREYVNARHCLPCHPECOPONGSVTCFGEADQCVACAHYKDP 594
Qy      601  FCVACSGGVKPDLSYMPIMKFPDEBGAQPCPINCTHSQVLDLXKCGPABEGRASPLTSI 660
Db      601  FCVACSGGVKPDLSYMPIMKFPDEBGAQPCPINCTHSQVLDLXKCGPABEGRASPLTSI 660
Qy      654  FCVACSGGVKPDLSYMPIMKFPDEBGAQPCPINCTHSQVLDLXKCGPABEGRASPLTSI 654
Db      654  FCVACSGGVKPDLSYMPIMKFPDEBGAQPCPINCTHSQVLDLXKCGPABEGRASPLTSI 654
Qy      720  VSAAVGIILLVVLGVVFGILIKRROOKIRKTYMRRLLOETLEVEPLTPSGAMPNOAMRI 720
Db      720  VSAAVGIILLVVLGVVFGILIKRROOKIRKTYMRRLLOETLEVEPLTPSGAMPNOAMRI 720
Qy      714  ISAVVGIILLVVLGVVFGILIKRROOKIRKTYMRRLLOETLEVEPLTPSGAMPNOAMRI 714
Db      714  ISAVVGIILLVVLGVVFGILIKRROOKIRKTYMRRLLOETLEVEPLTPSGAMPNOAMRI 714
Qy      781  LKETELARKVKGSGAFGTYYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 780
Db      781  LKETELARKVKGSGAFGTYYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 780
Qy      774  LKETELARKVKGSGAFGTYYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 774
Db      774  LKETELARKVKGSGAFGTYYKGIWIPDGENYKIPVAIKVLENTSPKANKEILDBAYVM 774
Qy      834  AGVSPYVSRLLGICLTSTVQLVTQLMPIYGLLDHVRENRGLSGODLLNMCQIAKGS 840
Db      834  AGVSPYVSRLLGICLTSTVQLVTQLMPIYGLLDHVRENRGLSGODLLNMCQIAKGS 840
Qy      841  YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGGKVPKMMALBSI 900
Db      841  YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGGKVPKMMALBSI 900
Qy      894  YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGGKVPKMMALBSI 894
Db      894  YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETRYHADGGKVPKMMALBSI 894
Qy      960  LRRRTTHOSDVMYSYGVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMI 960
Db      960  LRRRTTHOSDVMYSYGVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMI 960
Qy      954  LRRRTTHOSDVMYSYGVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMI 954
Db      954  LRRRTTHOSDVMYSYGVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYMI 954
Qy      1020  MVKCMIMIDSECRPRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSIFYNSLLEDDM 1020
Db      1020  MVKCMIMIDSECRPRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSIFYNSLLEDDM 1020
Qy      1014  MVKCMIMIDSECRPRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSIFYNSLLEDDM 1014
Db      1014  MVKCMIMIDSECRPRRELVSFESRMAARDPQRFVVIQNEIDLGPASPLDSIFYNSLLEDDM 1014
Qy      1080  GDVDAEAYLVPOQGFPCPDPAFGAGVWVHHRSSSTRSGGDLTLGLEPSEEAAPSP 1080
Db      1080  GDVDAEAYLVPOQGFPCPDPAFGAGVWVHHRSSSTRSGGDLTLGLEPSEEAAPSP 1080
Qy      1074  GDVDAEAYLVPOQGFPCPDPAFGAGVWVHHRSSSTRSGGDLTLGLEPSEEAAPSP 1074
Db      1074  GDVDAEAYLVPOQGFPCPDPAFGAGVWVHHRSSSTRSGGDLTLGLEPSEEAAPSP 1074
Qy      1140  LAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVALPICS 1140
Db      1140  LAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVALPICS 1140
Qy      1134  LAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVALPICS 1134
Db      1134  LAPSEGAGSDVFDGDLGMAKAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVALPICS 1134
Qy      1200  POPEYVNOVDYRPOPPSPREBGLPARAPAGATLEBAKTLSPGKNGVVDVAFGAVENP 1200
Db      1200  POPEYVNOVDYRPOPPSPREBGLPARAPAGATLEBAKTLSPGKNGVVDVAFGAVENP 1200

```

```

Db      1135  POPEYVNOVDYRPOPPSPREBGLPARAPAGATLEBAKTLSPGKNGVVDVAFGAVENP 1194
Qy      1201  EYLTPQGAAPQPHPPAPSPAFDNLYYWDOBPERRGAPSPSTFKGTPTAENPEYGLDVP 1260
Db      1195  EYLTPQGAAPQPHPPAPSPAFDNLYYWDOBPERRGAPSPSTFKGTPTAENPEYGLDVP 1254
Qy      1261  V 1261
Db      1255  V 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KM Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU ) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
XX
DR N-PSDB; AAO46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. Such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAO46083; AAR39568). X in AAR39568 represents
CC the location of a stop codon in AAO46083.
XX
SQ Sequence 1433 AA;

Query Match 96.2%; Score 6583; DB 14; Length 1433;
Best Local Similarity 96.5%; Pred. No. 0;
Matches 1217; Conservative 11; Mismatches 27; Indels 6; Gaps 1;

Qy      1  MELAALCRWGLLALIPGAASVQYCTGDMKRLPASPEHLDMRLYQCCQVYQGNL 60
Db      1  MELAALCRWGLLALIPGAASVQYCTGDMKRLPASPEHLDMRLYQCCQVYQGNL 60
Qy      61  ELYLPTNASLSFLDIOIEVQGVYLIAHNOVRQVPLQRLRIVRGTQLEFDNYALAVLNG 120
Db      61  ELYLPTNASLSFLDIOIEVQGVYLIAHNOVRQVPLQRLRIVRGTQLEFDNYALAVLNG 120
Qy      121  DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNOLA 180
Db      121  DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQRPOLCYOPTILMKDIFHKNOLA 180
Qy      181  LTLIDTNRSRACHPCSPMKCKSRGWGSESSDCCSLTRTVCAAGCARCKGPLPTDCHEQC 240
Db      181  LTLIDTNRSRACHPCSPMKCKSRGWGSESSDCCSLTRTVCAAGCARCKGPLPTDCHEQC 240
Qy      1261  V 1261
Db      1255  V 1255

```

QY 241 AAGCTGPHGSDCLACLFHNSGICELHCPALVTYNTDFESHPNDEGRYTFGASCVTACP 300
Db |
241 AAGCTGPHGSDCLACLFHNSGICELHCPALVTYNTDFESHPNDEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPLNHNOETAEADGTORCEKSKPCARVCYGLMOYIKANSKFIGIT 360
Db |
301 YNYLSTDVSGCTLVCPLNHNOETAEADGTORCEKSKPCARVCYGLMEHLREVAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLEPSEFDGDPASNTAPLQPEQLQVFETLEETGYLISAMPDSL 420
Db |
361 IOEFAGCKKIFGSLAFLEPSEFDGDPASNTAPLQPEHLQVFETLEQITGYLISAMPDSL 420
QY 421 DLSVFNQLQVIRGRILHNGAVSLTQGLGISWLGRLSRLRELGSLALIHNTLCEVHTV 480
Db |
421 DLSVFNQLQVIRGRILHNGAVSLTQGLGISWLGRLSRLRELGSLALIHNTLSFVHTV 480
QY 481 PMDOLFRNNFTVSWLVPKVSASHLEBCEVGEGLACHOLCARGHCWGPPTQCVNCSOF 540
Db |
481 PMDOLFRNPFOALHTANRPE----DECVGEGLACHOLCARGHCWGPPTQCVNCSOF 534
QY 541 LRGOECVECECVLQGLPREVYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHYKDP 600
Db |
535 LRGOECVECECVLQGLPREVYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHYKDP 594
QY 601 FCVABCPGCVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCCPAEQRASPLTSI 660
Db |
595 FCVABCPGCVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCCPAEQRASPLTSI 654
QY 661 USAVVGILVVLGVVFGILIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNOQMRI 720
Db |
655 ISAVVGILVVLGVVFGILIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEAYVM 780
Db |
715 LKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEAYVM 774
QY 781 AGVSPYVSRLLGICLTSTVOLVTOAMPYGLLDHVRENRGLSQDILLNMCQIAKMS 840
Db |
775 AGVSPYVSRLLGICLTSTVOLVTOAMPYGLLDHVRENRGLSQDILLNMCQIAKMS 834
QY 841 YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETEHADGGKVPKIMMALESI 900
Db |
835 YLEDEVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETEHADGGKVPKIMMALESI 894
QY 901 LRRRFTHQSDVMSYGVTVMEMLTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
Db |
895 LRRRFTHQSDVMSYGVTVMEMLTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
QY 961 MVKCMIDSECRPRPRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1020
Db |
955 MVKCMIDSECRPRPRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDM 1014
QY 1021 GDLVDABEYLVPOGGFFCPDPAPGAGGVHHRHSSSTRSGGDLTLGLEPSEBEARSP 1080
Db |
1015 GDLVDABEYLVPOGGFFCPDPAPGAGGVHHRHSSSTRSGGDLTLGLEPSEBEARSP 1074
QY 1081 LAPSEGAGSDVFDGDLGKAAGLQSLFTHDPSPLQRYSEDPVPLPSETDGYVAPLTC 1140
Db |
1075 LAPSEGAGSDVFDGDLGKAAGLQSLFTHDPSPLQRYSEDPVPLPSETDGYVAPLTC 1134
QY 1141 POPEYVNOQDVVRPOPSBREGPLPAARPAGATLERAKTLPKXGVVVDVAFGAVENP 1200
Db |
1135 POPEYVNOQDVVRPOPSBREGPLPAARPAGATLERAKTLPKXGVVVDVAFGAVENP 1194
QY 1201 EYLTPOGGAAPQPPAPFAPFNDLYWMDPPERGAPESTFKGTPTAENPEYLGLDVP 1260
Db |
1195 EYLTPOGGAAPQPPAPFAPFNDLYWMDPPERGAPESTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

Search completed: July 22, 2003, 08:41:40
Jcb time : 44.1589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 6841

Sequence: 1 MELALCRMGLLALLPPGA.....TFKGTPTANPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6621	96.8	1255	1 A24571	protein-tyrosine k
2	5827	85.2	1260	1 TVR1NU	protein-tyrosine k
3	5817.5	85.0	1254	2 I48161	p-185 precursor -
4	3110	45.5	1210	1 GOUHE	epidermal growth f
5	3079	45.0	1210	2 A53183	epidermal growth f
6	3053.5	44.6	1223	1 TVCHLV	epidermal growth f
7	2945.5	43.1	1308	2 A47253	epidermal growth f
8	2627	38.4	1166	1 S06142	epidermal growth f
9	2378.5	34.8	1342	2 A36223	kinase-related tra
10	2293.5	33.5	1339	2 JCA387	epidermal growth f
11	1725.5	25.2	698	1 TVPVLV	protein-tyrosine k
12	1677	24.5	604	1 TVTTH	protein-tyrosine k
13	1625.5	23.8	1330	1 GQFE	epidermal growth f
14	1606	23.5	544	2 S03745	epidermal growth f
15	1599	23.4	545	2 S00727	protein-tyrosine k
16	1582	23.1	540	2 B44776	kinase-related tra
17	1580	23.1	540	1 TVPVEB	protein-tyrosine k
18	1464	21.4	644	2 A36225	epidermal growth f
19	1305	19.1	1223	2 E88257	epidermal growth f
20	1305	19.1	1374	2 S70712	protein-tyrosine k
21	1220	17.8	1369	2 S70713	protein-tyrosine k
22	1176	17.2	1717	1 A45558	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	979.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	730	10.7	1363	2 T43220	insulin-like growth
28	696	10.2	1382	1 INHUR	insulin receptor p
29	694	10.1	1383	2 A36080	insulin receptor p

30	693	10.1	1477	2 T18534	protein-tyrosine k
31	691.5	10.1	1372	2 A34157	insulin receptor p
32	691	10.1	1607	2 T43212	insulin-like growth
33	687.5	10.0	1300	2 A36502	insulin receptor-x
34	678	9.9	1268	2 B36502	insulin receptor-x
35	647	9.5	1367	1 IGHUR1	insulin-like growth
36	628	9.2	1371	2 A33837	insulin-like growth
37	620.5	9.1	2148	1 A56081	insulin receptor -
38	620	9.1	1390	2 T30346	insulin receptor -
39	614	9.0	2101	2 S57245	insulin receptor (
40	597	8.7	987	2 A54092	protein-tyrosine k
41	594.5	8.7	977	2 S49004	tyrosine kinase Mp
42	591.5	8.6	984	2 A39753	protein-tyrosine k
43	588.5	8.6	976	2 A36355	protein-tyrosine k
44	586.5	8.6	952	2 I50612	protein-tyrosine k
45	586	8.6	1091	2 S33596	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein erl

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.

Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human C-erb-B-2 gene to epidermal growth

A/Reference number: A24571; PMID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 11255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, C-erbB-2, is distinct from the C-erbB-1/epide:

A/Reference number: A25491; PMID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1137-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chrom

A/Reference number: A44188; PMID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Residues: 740-910 <COU1>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; NID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:G459607; PIDN:AAA35809.1; PID:G459608

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcription;

A/Reference number: I57622; PMID:8728698; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:g183993; PIDN:AA58637.1; PID:g553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics: GDB:ERBB2; NGL; NEU; HER-2
 A:Gene: GDB:ERBB2; NGL; NEU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-155/Produce: protein-tyrosine kinase erbb2 #status predicted <MAT>
 F:22-653/Domain: extracellular #status predicted <EXT>
 F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F:654-675/Domain: transmembrane #status predicted <TM>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68,124,187,259,530,571,629/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.8%; Score 6621; DB 1; Length 1255;
 Best Local Similarity 96.8%; Pred. No. 2,66-267;
 Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

```

1  MELAALCRMGGLALLALPPGAASSTOVCTGTMKRLRPASPTHMLMLRHLVYGCQVQGNL 60
Db 1  MELAALCRMGGLALLALPPGAASSTOVCTGTMKRLRPASPTHMLMLRHLVYGCQVQGNL 60
Qy 61  ELTYLPTNASLSPLODIOEVGVYVLAHNOVQVPLQRLIRVSTQDFEDNYALAVLDNG 120
Db 61  ELTYLPTNASLSPLODIOEVGVYVLAHNOVQVPLQRLIRVSTQDFEDNYALAVLDNG 120
Qy 121  DPLNNTPTVVGASGGLREQLRSLTEILKGVLIQNPOLCYODTILMDIFHKNQOLA 180
Db 121  DPLNNTPTVVGASGGLREQLRSLTEILKGVLIQNPOLCYODTILMDIFHKNQOLA 180
Qy 181  LTLIDTRSRACHPCSPMCKSGRSCGSSSEDCSLTFTVCAGGACRCKGPIPTDCCHEQC 240
Db 181  LTLIDTRSRACHPCSPMCKSGRSCGSSSEDCSLTFTVCAGGACRCKGPIPTDCCHEQC 240
Qy 241  AAGCTGPKHSDCLALCFHNSGICELHCPALVTYNTDFTESMPNBEGRYTFGASCVTACP 300
Db 241  AAGCTGPKHSDCLALCFHNSGICELHCPALVTYNTDFTESMPNBEGRYTFGASCVTACP 300
Qy 301  YNLTSTDVGSCTLVCPRLHNOEVTAEDEGTQREKSKCARVCYGLGMQYIKANSKPIGIT 360
Db 301  YNLTSTDVGSCTLVCPRLHNOEVTAEDEGTQREKSKCARVCYGLGMQYIKANSKPIGIT 360
Qy 361  ELEFAGCKKIFGSLAFPESEFGDPASNTAPLOPEOVFFTEITGYLYISAMPDLP 420
Db 361  IOEFAGCKKIFGSLAFPESEFGDPASNTAPLOPEOVFFTEITGYLYISAMPDLP 420
Qy 421  DLSVFQNLQVIRGRILHNGAVSLTLQGLISWLGRLSLRELGSGLAIHNNTHLCFVHTV 480
Db 421  DLSVFQNLQVIRGRILHNGAVSLTLQGLISWLGRLSLRELGSGLAIHNNTHLCFVHTV 480
Qy 481  PMOQLPNNPHOALLHTANRPEDECVGEGACHOICAGGHCMPPTQCVNCSQSLRQDEC 540
Db 481  PMOQLPNNPHOALLHTANRPEDECVGEGACHOICAGGHCMPPTQCVNCSQSLRQDEC 540
Qy 541  VEEGRVLQGLPREYVNAHRLCPHRECOPOGNGSTCGPEADQCVACAHYDPPFCVARC 600
Db 541  VEEGRVLQGLPREYVNAHRLCPHRECOPOGNGSTCGPEADQCVACAHYDPPFCVARC 600
Qy 601  PFNNFTVSPWLRVFKVSASHLE---PDEGACOPPCPINTGSCVLDLDDKCCPAEQRASP 656
Db 601  PFNNFTVSPWLRVFKVSASHLE---PDEGACOPPCPINTGSCVLDLDDKCCPAEQRASP 656

```

Db 601 PSG-----VKPDLSTYMPIMKPDDEGACOPPCPINTGSCVLDLDDKCCPAEQRASP 650
 Qy 657 LTSIVSAVNGILLVNVLGVPFGLIKRQOKIRKTYTRRLLOQELVELPELTPSGAMPQA 716
 Db 651 LTSIVSAVNGILLVNVLGVPFGLIKRQOKIRKTYTRRLLOQELVELPELTPSGAMPQA 710
 Qy 717 QMRILKETEIRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKAKKELIDE 776
 Db 711 QMRILKETEIRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKAKKELIDE 770
 Qy 777 AYMAGVSPYVSRLLGICLTSTVQVLTQMPYGCCLLDHYRENRGRLSQDLNMCQIA 836
 Db 771 AYMAGVSPYVSRLLGICLTSTVQVLTQMPYGCCLLDHYRENRGRLSQDLNMCQIA 830
 Qy 837 KGMSTLEDVNLVNRDLAARVLYKSPNNVKTTPDEGLARLLDIDETEHADGKPIKMA 896
 Db 831 KGMSTLEDVNLVNRDLAARVLYKSPNNVKTTPDEGLARLLDIDETEHADGKPIKMA 890
 Qy 897 LESTLRFPFHOSDWSYGVTVMEIMTFGAKPYGIPAREIPDLLEKGERLPOPICTID 956
 Db 891 LESTLRFPFHOSDWSYGVTVMEIMTFGAKPYGIPAREIPDLLEKGERLPOPICTID 950
 Qy 957 VYMIMVCMNIDSECRPFRELVSSEPSRMARDPQRFVIONEDLGAPSPDSTFYRLLE 1016
 Db 951 VYMIMVCMNIDSECRPFRELVSSEPSRMARDPQRFVIONEDLGAPSPDSTFYRLLE 1010
 Qy 1017 DDDMGDLVDABEYLVPQGGFPCDPAPGAGGMYHRRSSSTRSGGDLTLGLEPSEBEA 1076
 Db 1011 DDDMGDLVDABEYLVPQGGFPCDPAPGAGGMYHRRSSSTRSGGDLTLGLEPSEBEA 1070
 Qy 1077 PRSPPLASBEGASDVPFGDGLGMAKGLQSLPTDPSPLORYSDDPTVPLPSETDGYAP 1136
 Db 1071 PRSPPLASBEGASDVPFGDGLGMAKGLQSLPTDPSPLORYSDDPTVPLPSETDGYAP 1130
 Qy 1137 LTCSPQPEYVNPQDVRPQPSPREGPLPAARPAATLERAKTSLPGKNGVYKVFAPGCA 1196
 Db 1131 LTCSPQPEYVNPQDVRPQPSPREGPLPAARPAATLERAKTSLPGKNGVYKVFAPGCA 1190
 Qy 1197 VENPEYLTPOGGAAPQHPPPAPSPAFDNLTYMDQDPEPKAPSTKGTATAPPEYL 1256
 Db 1191 VENPEYLTPOGGAAPQHPPPAPSPAFDNLTYMDQDPEPKAPSTKGTATAPPEYL 1250
 Qy 1257 LDVPV 1261
 Db 1251 LDVPV 1255

RESULT 2
 TVRTNU
 A:protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1988 #sequence-revision 31-Dec-1988 #text-change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Barbmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
 A:Reference number: A24562; MUID:86118662; PMID:3945311
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 1-1260 <BAR>
 R:Mani, T.; Mann, A.M.; McGarree, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen
 Carcinogenesis 12, 1975-1978, 1991
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
 2-thiathiolylformamide or N-methyl-N-iltrosourea.
 A:Reference number: A61204; MUID:92035293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics:
 A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAM>
 F:658-680/Domain: transmembrane #status predicted <TMN>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:711,791,263,535,576,634/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: Lys #status predicted
 F:882,1227/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5827; DB 1; Length 1260;
 Best Local Similarity 85.3%; Pred. No. 2e-234;
 Matches 1081; Conservative 53; Mismatches 117; Indels 16; Gaps 4;

```

OY 1 MELAALCRMGILLALALPPGAASTQVCTGDMKRLPASPEHLDMLRHLYOGCQVQGNL 60
DB 4 MELAAMCRMGILLALALPPGAGTQVCTGDMKRLPASPEHLDMLRHLYOGCQVQGNL 63
OY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVQVPLQRLRIYRGTOQLFEDNYALAVLDNG 120
DB 64 ELTYVPANASLSFLQDIOEVQGVYLAHNOVQVPLQRLRIYRGTOQLFEDNYALAVLDNR 123
OY 121 DPLANTTPVT-GASPGGLRELQRLSLTEILKGVLLQRNQOLCYODTIIMKDIIFHKNNOL 179
DB 124 DPLDNTTPVTGRTPGRLQLRLSLTEILKGVLLQRNQOLCYODTIIMKDIIFHKNNOL 183
OY 180 ALTLIDTNSRACHPCSPMKSGRCGSESEDCOSLTRITACAGGACRCKGLPTDCCHEQ 239
DB 184 APVDIDTNSRACHPCSPACKNDHNCWGESPEDCOLITGITTCGACRCKGLPTDCCHEQ 243
OY 240 CAAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRTYFGASCVTAC 299
DB 244 CAAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRTYFGASCVTTC 303
OY 300 PNYLSTDVGSCTLVPLRNOETADGTCRCCKSPCARVCYGIAMQYIKANSKPIGT 359
DB 304 PNYLSTDVGSCTLVPLRNOETADGTCRCCKSPCARVCYGIAMQYIKANSKPIGT 363
OY 360 TELEFAGCKKIFGSLAFLESFPGDPAASNTAPLPOLQVFEFLTEITGVLYISAMPDLS 419
DB 364 NVOEPGCKKIFGSLAFLESFPGDPAASNTAPLPOLQVFEFLTEITGVLYISAMPDLS 423
OY 420 PDLVSFQNLQVIRGRILHNGAYSLSLQGLISWLGRLSRELSGLALIHNNHLCFVHT 479
DB 424 PDLVSFQNLQVIRGRILHNGAYSLSLQGLISWLGRLSRELSGLALIHNNHLCFVHT 483
OY 480 VPMQDLFRPHQALLHTARPEDE-CVGGGLACHQULCANGHCKGPGPTCCVNCSGFLRQ 538
DB 484 VPMQDLFRPHQALLHTARPEDE-CVGGGLACHQULCANGHCKGPGPTCCVNCSGFLRQ 543
OY 539 ECVEECRVLQGLPREVYNARHCLPCHREGOPONGSVCFCGRPADOCVACAHYNDPFCVA 598
DB 544 ECVEECRVLQGLPREVYNARHCLPCHREGOPONGSVCFCGRPADOCVACAHYNDPFCVA 603
OY 599 RCPENNFTVSPMLRVKVSASHLE---PDEBGCOPCPDINCHTSCVLDLDDKCPAEPORA 654
DB 604 RCPENFTVSPMLRVKVSASHLE---PDEBGCOPCPDINCHTSCVLDLDDKCPAEPORA 653
OY 655 SPLTISVAVNGILLVNVGVVFGILLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPN 714
DB 654 SPLTISVAVNGILLVNVGVVFGILLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPN 713
OY 715 QAKMRLKETELRKVYVLSGSAFGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKETL 774
DB 714 QAKMRLKETELRKVYVLSGSAFGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKETL 773
OY 775 DEAYMAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYRENGRLGSDLLMNCQ 834
DB 774 DEAYMAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYRENGRLGSDLLMNCQ 833
OY 835 IAKGMSYLEDVVLVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETETHADGKVPKIM 894

```

```

DB 834 IAKGMSYLEDVVLVHRDLAARNVLYKSPNHVKITDPGLARLLDIDETETHADGKVPKIM 893
OY 895 MALESLIRRRFTQSDVMSYGVVWMLMTGAPYDGI PAREIPDLEKGRLLPOPICT 954
DB 894 MALESLIRRRFTQSDVMSYGVVWMLMTGAPYDGI PAREIPDLEKGRLLPOPICT 953
OY 955 IDVYIMVWKCMIDSECRPFRELVEFSFMAADPQRFVYIIONEDLGPASPLDSTFYRSL 1014
DB 954 IDVYIMVWKCMIDSECRPFRELVEFSFMAADPQRFVYIIONEDLGPASPLDSTFYRSL 1013
OY 1015 LEDDDMGDLVDAEYLVPOCGFFCPDPAPAGGVHHRHRSSTSGCGDLTLCLEPSEE 1074
DB 1014 LEDDDMGDLVDAEYLVPOCGFFCPDPAPAGGVHHRHRSSTSGCGDLTLCLEPSEE 1073
OY 1075 EAPRSLPASEGASGVFDDDLGMGAAGKQSLPTHDPSLQRYSEPTVPLPSETGVY 1134
DB 1074 GPRSLPASEGASGVFDDDLGMGAAGKQSLPTHDPSLQRYSEPTVPLPSETGVY 1133
OY 1135 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAAPAGATLEBRAKTLSPGKGVVADVAFG 1194
DB 1134 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAAPAGATLEBRAKTLSPGKGVVADVAFG 1193
OY 1195 GAVENPEYLVTPQCGAALPQHPPPAFSPAFDNLVYQDPPREGAPSTFKGTPTAENPEY 1254
DB 1194 GAVENPEYLVTPQCGAALPQHPPPAFSPAFDNLVYQDPPREGAPSTFKGTPTAENPEY 1253
OY 1255 LGLDVPV 1261
DB 1254 LGLDVPV 1260

```

RESULT 3
 148161
 p-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_rev150 02-Jul-1996 #text_change 18-Jun-1999
 C:Accession: 148161
 R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
 Gene 140: 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: 148161; MID:94193007; PMID:7908275
 A:Accession: 148161
 A:Molecule type: mRNA
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Residues: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:9493236; PIDN:BA03801.1; PID:9747595
 C:Gene: neu
 A:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:718-983/Domain: protein kinase ATP-binding motif
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.0%; Score 5817.5; DB 2; Length 1254;
 Best Local Similarity 85.0%; Pred. No. 5e-234;
 Matches 1075; Conservative 62; Mismatches 113; Indels 15; Gaps 3;

```

OY 1 MELAALCRMGILLALALPPGAASTQVCTGDMKRLPASPEHLDMLRHLYOGCQVQGNL 60
DB 1 MELAAMCGMGLLALALPPGASGTQVCTGDMKRLPASPEHLDMLRHLYOGCQVQGNL 60
OY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVQVPLQRLRIYRGTOQLFEDNYALAVLDNG 120
DB 61 ELTYVPANASLSFLQDIOEVQGVYLAHNOVQVPLQRLRIYRGTOQLFEDNYALAVLDNR 120
OY 121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLLQRNQOLCYODTIIMKDIIFHKNNOL 180
DB 121 DPLANTTPVTGASPGGLRELQRLSLTEILKGVLLQRNQOLCYODTIIMKDIIFHKNNOL 180
OY 180 ALTLIDTNSRACHPCSPMKSGRCGSESEDCOSLTRITACAGGACRCKGLPTDCCHEQ 240
DB 181 PVDIDTNSRACHPCSPACKNDHNCWGESPEDCOLITGITAPRAVPAARARLLPTDCCHEQ 240

```

```

QY 241 AAGCTGKSHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTGASCATYACP 300
DB 241 AAGCTGKSHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTGASCATYACP 300
QY 301 YNYLSTVSGCTTLCPLHNOEVTADGTORCEKSKPCARCYGCMQYIKANSKFTIGIT 360
DB 301 YNYLSTVSGCTTLCPLHNOEVTADGTORCEKSKPCARCYGCMQYIKANSKFTIGIT 360
QY 361 ELEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDLP 420
DB 361 IOEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDLP 420
QY 421 DLSEFOGLVYRGRILNNGAYSLTLOGIGLWMLGRSLRELGSGLVLIHNRTHLCFVHTV 480
DB 421 DLSEFOGLVYRGRILNNGAYSLTLOGIGLWMLGRSLRELGSGLVLIHNRTHLCFVHTV 480
QY 481 PMDQLFNNPQALHTANRPEDECVCGEGLACHOLCARGHCWGPPTOCVNCQFLAQEC 540
DB 481 PMDQLFNNPQALHTANRPEDECVCGEGLACHOLCARGHCWGPPTOCVNCQFLAQEC 540
QY 541 VEEBRYVGLPREYVYVNAHCLPCHECOPONGSVTCGPEADQCAHYKDPPECVARC 600
DB 541 VKECRVWKGLPREYVYVNAHCLPCHECOPONGSVTCGPEADQCAHYKDPPECVARC 600
QY 601 PFNNFTVSFMLRVPKVSASHLE----PDEGACQPCPINCNCVLDLDDKCPAEORASP 656
DB 601 PSC-----VKPDLSTWPIWKYPRBEMCMCPCLNCHSCVDLDERCPEAEORASP 650
QY 657 LTSIVSAVGLLVVVLGVVFGILIKRQOKIRKRYTRRLLOETLEVEPLTPSGAMPNOA 716
DB 651 ATSIIVAVGILFLVIGVVVGLIKRRQKIRKRYTRRLLOETLEVEPLTPSGAMPNOA 710
QY 717 QMRILKTELKRVAVLSSGAGTYKGIWIPDGENVKIPVAKYLRENTSKANKELIDE 776
DB 711 QMRILKTELKRVAVLSSGAGTYKGIWIPDGENVKIPVAKYLRENTSKANKELIDE 770
QY 777 AYWAGSGPVSVSLGLCLTSTVQLVTQLMPYGCCLDHYVENGRGLSODLLMCMQOIA 836
DB 771 AYWAGSGPVSVSLGLCLTSTVQLVTQLMPYGCCLDHYVENGRGLSODLLMCMQOIA 830
QY 837 KGSMSYLEDVRLVHRDLAARNVLYKSPNNVKITDFGLARLLDIDETEVHADGKVPYIKMA 896
DB 831 KGSMSYLEDVRLVHRDLAARNVLYKSPNNVKITDFGLARLLDIDETEVHADGKVPYIKMA 890
QY 897 LESLIRRRFTHQSVMSVGYVWELMTFGARPYGCIYAREIPDLLEKGERLPPOPICTID 956
DB 891 LESLIRRRFTHQSVMSVGYVWELMTFGARPYGCIYAREIPDLLEKGERLPPOPICTID 950
QY 957 VYMIWVCMWIDSECRPRELVESEFSRMARDPQRFVYIQNEDLGPASPILDSTFYRSLLE 1016
DB 951 VYMIWVCMWIDSECRPRELVESEFSRMARDPQRFVYIQNEDLGPASPILDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDPAQAGVMVHRRSSSTRSGGDLTLGLEPSEBEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDPAQAGVMVHRRSSSTRSGGDLTLGLEPSEBEA 1070
QY 1077 PRSGLASBEGAGSVPCDLCMGAKGLQSLPTHDPSPLOQYSSDPVYPLPSEFDGYAP 1136
DB 1071 PRSGLASBEGAGSVPCDLCMGAKGLQSLPTHDPSPLOQYSSDPVYPLPSEFDGYAP 1130
QY 1137 LTCSPOPEYVNOQVRRPQPSRREGPLPAARPAAGATLERAKTSLPGKNGVYKDFARFGA 1196
DB 1131 LACSPOPEYVNOQVRRPQPSRREGPLPAARPAAGATLERAKTSLPGKNGVYKDFARFGA 1190
QY 1197 VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDODPPERGAPSTKGTPTAENPEYLG 1256
DB 1191 VENPEYLVPRGSGASQPH-PPALCPARDNLVYWDODPPERGAPSTKGTPTAENPEYLG 1249
QY 1257 LDVPU 1261
DB 1250 LDVPU 1254

```

RESULT 4

```

GQHE
Epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erdb
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
A:Accession: A00641, A25772, S30024, A38672, A0642, A43615, A23062, A05281, A60143, A
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641, MUID:84219729, PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <UL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA52240.1; PID:g757924
A>Note: the authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal g
A:Reference number: A25772, MUID:85270438, PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M1234; NID:g181991; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kitchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A:Reference number: S30024, MUID:8821333, PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termin
A:Reference number: A38672, MUID:91107677, PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
R:Xu, Y.; Ishii, S.; Clark, A.O.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; M
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNA
A:Reference number: A00642, MUID:84245835, PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 1-187, 'KSVYQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-3
', 798-799, 'TD', 802-811, 'R', 813-942 <XUV>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF rec
R:Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificatio
A:Reference number: A43615, MUID:84196372, PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062, MUID:85046483, PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Meber, W.; Gill, G.N.; Speiser, J.
Science 224, 294-297, 1984
A:Reference number: A05281, MUID:84172183, PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

```


R.Russo, M.W.; Lukac, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A:Reference number: A60143; MUID:85182650; PMID:2985580
 A:Accession: A60143
 A:Molecule type: protein
 A:Residues: 740-744, X, 746-747 <RUS>
 R:Motoczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A:Reference number: A38023; MUID:84191554; PMID:6325948
 A:Contents: annotation, receptor activity
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.; Cell 59, 33-43, 1989
 A:Title: Functional independence of the epidermal growth factor receptor from a domain A:Reference number: A3331; MUID:90002233; PMID:2790960
 A:Contents: annotation, internalization signal
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C:Genetics:
 A:Gene: GDB:EGFR
 A:Cross-references: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase; signal sequence #status predicted <SIG>
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1210/Product: EGF receptor #status predicted <MAT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:1390-300/Domain: EGF receptor extracellular domain repeat <EET>
 F:1390-600/Domain: EGF receptor extracellular domain repeat <EET>
 F:446-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:745/Active site: Lys #status experimental

Query Match 45.54; Score 3110; DB 1; Length 1210;
 Best local Similarity 49.64; Pred. No. 7, 6e-122; Indels 124; Gaps 26;
 Matches 633; Conservative 170; Mismatches 350;

```

11 LLLALLPPGAA--STOVCTGTMKRLPASPTHLDMLRYOCQVYQGNLEITYLPTN 68
14 LLAALCPASRALREKVKCGCTSNKLTQLGFEDHPLSLQMFNNECVLNLLETYYQRN 73
69 ASLSFLODIQVGVYLIANQVROVPLQRLIVRGTLQFEDNYALAVLDNGPLNNTTP 128
74 YDLSFLKTIQEVGVYLIANLTVTERIPLENLQIRGMVYENSVALAVLSNYD----- 126
129 VTGASFGGLRELDRLSITELKGVLIQRPOLCYODTILMKDIFHNQNLATLLIDTR 168
127 ---ANKTGLKELPMRLQDELHGAVRFSSNPALCNVESIQMRDVISSDFLSNMKDFQ 183
189 SPACHPCSPWCKSGSRGSSSEDQSLTRTVCAAGCA-RCKGPLPTDCHEQCAAGTGP 247
184 LGSQCKDCSPGSCMGAGEENCKLTKIICAQCSGRCRGSFSCCHNQCAAGTGP 243
248 KASDCLACHFNHSGICELHCPALVYNTDFESMRPERRYFGASCCTACPYNTLSD 307
244 RSDCLVCRKFRDEATCKTCTCPPLMYNPTTYQMDVNPBEKVSFGATCVKCCPRNVYTD 303
308 VSSCTLVCPHNOEVAEDGTORCEKSPKCAVCGYGLGMQYIKANSKFIQTELE-FAG 366
304 HSCSVACGADSVEM-EEDGVKRCCKCEGCRKVCIGIGERK-DLSLNATINIKHFKN 361
367 CKKIFGSLAFLPESFGDPASNTAPLQPEQLQVFEFLLEETLYLYISAMPDLSPLDVFQ 426
362 CTSISGDLHLPAFAFGDSFTTPPLDPOBLDKTKVKEITGFLLIQAWPENNTDLHAF 421
427 NLOVGRGLNHGAVSLTQGLGISWLGLSLRELQSLGLLIHNHNLCAVHTVPDQLF 486

```

```

Db 422 NLEIIRRTYQHGFPSLAIVSLNITSGLSLKEISDGVIIISGNKMLCYANTINMKKL 481
Db 487 RNPQALLHTNRPDECEVSGELACHOLCARGHCGPPTQCVNCSQFLRGCEVEECRV 546
Db 482 GTSQKTKIISNRKENSCKATGVCHALCSPEGCGEPFRCVCSNRVSRGECVDCKL 541
Db 547 LQGLPREYVNAHRLCPHRECOPQNSVTCFGEPAQOCVACAHKDPFCVACRPF----- 602
Db 542 LEGSPREFEVESECTICQHPECLPQAMNITCTGRBPNDICQAHIDGPHCVKTCGAVWG 601
Db 603 -NNTVTFYLRVPKVSASHLEDEGACOPFPICTHSVDLDDKCGCAEQASPLTSIV 661
Db 602 ENNTLV--W--KYADAGH-----CHLCHPACTGCTGPGEGCPTNGPKIP--SIA 647
Db 662 SAUVG---ILLVVLGVVFGILIRROOKIRKYMRLQETTELVEPTPGSCAMNQOM 718
Db 648 TGMVGLALLLVVVLGIG--LFRKRHHYAKRTLRLRLQRELVEPLTPGEARNQLL 704
Db 719 RLKETELRKVKVLSGAFVTYKGIWIPDENYKIPVAIKVLRENTS PKANKETLDEAY 778
Db 705 RLKETEFKKIKVLSGAFVTYKGLWIPDEYKIPVAIKELKEATS PKANKETLDEAY 764
Db 779 VMAGVSPYVSLIGLITLTVQLVTLQMPYGLLDHYRENRGLSGODLLNCKQIAKG 838
Db 765 VMASVDNPHVCRLLGICLTSTVQLITQIMPGLLDYREKDNIGSQVLLNMCVQIAKG 824
Db 839 MSYLEDVRLVHRDLAARNVLYKSPNHYKINDPGLARLLIDETETHADGCVPIKMALE 898
Db 825 KMYLEDRLVHRDLAARNVLYKTPQHVKTIDPGLAKLGAEBEYVHAEGVPIKMALE 884
Db 899 SILRRRTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPOPPICTIDVY 958
Db 885 SILHRIYTHOSDWSYGVTVWELMTFGSKPYDGIPIAREISILKGRLLPOPPICTIDVY 944
Db 959 MIMVCKMIDSECRPRRELVEFSRMAADPQRFVITQ-NEDLCPASPLDSTFYRSLLED 1017
Db 945 MIMVCKMIDADSPKFERELIEFSKMAADPQRFVITQ-NEDEKPMPLPSPTSNFPYALMDE 1004
Db 1018 DDMGDLVDAREEYLVPQCGFCPPDAPAGCMVHHRHSSSTRSGGDLTLGFESEEAR 1077
Db 1005 EDMDDVDVDADEYLLPQCGFF-----SSPSTS 1030
Db 1078 RSLPABEGAGSDVDFDGLGMAAGLQSLPTHDPSPLOYRSDPTVLPSET--DGVA 1135
Db 1031 RTPLSLSLATSN--NSTVACIDRNGLOSCPIKEDSLQKYSPTGALTEDSIDDTFL- 1087
Db 1136 PLTCSQPEVYNQDVARPPSPRECEPLPARAPAGATLERAKTILSPKNGVAVXDFAFGG 1195
Db 1088 -----PVPEYING-SVPEKPAQSVQNVYHNPQPLNP-----AFSRDPHYQD--PHST 1131
Db 1196 AVENPEYL--TPQGAAPQHPPPAPFAPFNLVYWDQ-----DP-----PERGA 1238
Db 1132 AVGNPEYLVQ-----PTCVNSTFDSPAHMAQKSHQISLDNDVQODFFPREAK 1162
Db 1239 PPSTFGKTPAENPEYL 1255
Db 1183 PNGIFKGS-TAENAEYL 1198

```

RESULT 5
 A53183
 Epidermal growth factor receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C:Accession: A53183; A43818; S24942; A28941; S45325; I49643
 R:Luetjens, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A. Genes Dev. 8, 399-413, 1994
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
 A:Reference number: A53183; MUID:9410986; PMID:8125255
 A:Accession: A53183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <LUS>
 A:Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A:Reference number: A43818; MUID:91232866; PMID:2030916
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Blisinger, D.P.; Serrero, G.
Submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
R:Heiseremann, G.V.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: Protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009, 'R', 1100-1101, 'L', 1103-1104, 'A', 1106-1107, 'S', 1109-1110, 'S', 1112-1113, 'S', 1115-1116, 'S', 1118-1119, 'S', 1121-1122, 'S', 1124-1125, 'S', 1127-1128, 'S', 1130-1131, 'S', 1133-1134, 'S', 1136-1137, 'S', 1139-1140, 'S', 1142-1143, 'S', 1145-1146, 'S', 1148-1149, 'S', 1151-1152, 'S', 1154-1155, 'S', 1157-1158, 'S', 1160-1161, 'S', 1163-1164, 'S', 1166-1167, 'S', 1169-1170, 'S', 1172-1173, 'S', 1175-1176, 'S', 1178-1179, 'S', 1181-1182, 'S', 1184-1185, 'S', 1187-1188, 'S', 1190-1191, 'S', 1193-1194, 'S', 1196-1197, 'S', 1199-1200, 'S', 1202-1203, 'S', 1205-1206, 'S', 1208-1209, 'S', 1211-1212, 'S', 1214-1215, 'S', 1217-1218, 'S', 1220-1221, 'S', 1223-1224, 'S', 1226-1227, 'S', 1229-1230, 'S', 1232-1233, 'S', 1235-1236, 'S', 1238-1239, 'S', 1241-1242, 'S', 1244-1245, 'S', 1247-1248, 'S', 1250-1251, 'S', 1253-1254, 'S', 1256-1257, 'S', 1259-1260, 'S', 1262-1263, 'S', 1265-1266, 'S', 1268-1269, 'S', 1271-1272, 'S', 1273-1274, 'S', 1276-1277, 'S', 1279-1280, 'S', 1282-1283, 'S', 1285-1286, 'S', 1288-1289, 'S', 1291-1292, 'S', 1294-1295, 'S', 1297-1298, 'S', 1300-1301, 'S', 1303-1304, 'S', 1306-1307, 'S', 1309-1310, 'S', 1312-1313, 'S', 1315-1316, 'S', 1318-1319, 'S', 1321-1322, 'S', 1324-1325, 'S', 1327-1328, 'S', 1330-1331, 'S', 1333-1334, 'S', 1336-1337, 'S', 1339-1340, 'S', 1342-1343, 'S', 1345-1346, 'S', 1348-1349, 'S', 1351-1352, 'S', 1354-1355, 'S', 1357-1358, 'S', 1360-1361, 'S', 1363-1364, 'S', 1366-1367, 'S', 1369-1370, 'S', 1372-1373, 'S', 1375-1376, 'S', 1378-1379, 'S', 1381-1382, 'S', 1384-1385, 'S', 1387-1388, 'S', 1390-1391, 'S', 1393-1394, 'S', 1396-1397, 'S', 1399-1400, 'S', 1402-1403, 'S', 1405-1406, 'S', 1408-1409, 'S', 1411-1412, 'S', 1414-1415, 'S', 1417-1418, 'S', 1420-1421, 'S', 1423-1424, 'S', 1426-1427, 'S', 1429-1430, 'S', 1432-1433, 'S', 1435-1436, 'S', 1438-1439, 'S', 1441-1442, 'S', 1444-1445, 'S', 1447-1448, 'S', 1450-1451, 'S', 1453-1454, 'S', 1456-1457, 'S', 1459-1460, 'S', 1462-1463, 'S', 1465-1466, 'S', 1468-1469, 'S', 1471-1472, 'S', 1474-1475, 'S', 1477-1478, 'S', 1480-1481, 'S', 1483-1484, 'S', 1486-1487, 'S', 1489-1490, 'S', 1492-1493, 'S', 1495-1496, 'S', 1498-1499, 'S', 1501-1502, 'S', 1504-1505, 'S', 1507-1508, 'S', 1510-1511, 'S', 1513-1514, 'S', 1516-1517, 'S', 1519-1520, 'S', 1522-1523, 'S', 1525-1526, 'S', 1528-1529, 'S', 1531-1532, 'S', 1534-1535, 'S', 1537-1538, 'S', 1540-1541, 'S', 1543-1544, 'S', 1546-1547, 'S', 1549-1550, 'S', 1552-1553, 'S', 1555-1556, 'S', 1558-1559, 'S', 1561-1562, 'S', 1564-1565, 'S', 1567-1568, 'S', 1570-1571, 'S', 1573-1574, 'S', 1576-1577, 'S', 1579-1580, 'S', 1582-1583, 'S', 1585-1586, 'S', 1588-1589, 'S', 1591-1592, 'S', 1594-1595, 'S', 1597-1598, 'S', 1600-1601, 'S', 1603-1604, 'S', 1606-1607, 'S', 1609-1610, 'S', 1612-1613, 'S', 1615-1616, 'S', 1618-1619, 'S', 1621-1622, 'S', 1624-1625, 'S', 1627-1628, 'S', 1630-1631, 'S', 1633-1634, 'S', 1636-1637, 'S', 1639-1640, 'S', 1642-1643, 'S', 1645-1646, 'S', 1648-1649, 'S', 1651-1652, 'S', 1654-1655, 'S', 1657-1658, 'S', 1660-1661, 'S', 1663-1664, 'S', 1666-1667, 'S', 1669-1670, 'S', 1672-1673, 'S', 1675-1676, 'S', 1678-1679, 'S', 1681-1682, 'S', 1684-1685, 'S', 1687-1688, 'S', 1690-1691, 'S', 1693-1694, 'S', 1696-1697, 'S', 1699-1700, 'S', 1702-1703, 'S', 1705-1706, 'S', 1708-1709, 'S', 1711-1712, 'S', 1714-1715, 'S', 1717-1718, 'S', 1720-1721, 'S', 1723-1724, 'S', 1726-1727, 'S', 1729-1730, 'S', 1732-1733, 'S', 1735-1736, 'S', 1738-1739, 'S', 1741-1742, 'S', 1744-1745, 'S', 1747-1748, 'S', 1750-1751, 'S', 1753-1754, 'S', 1756-1757, 'S', 1759-1760, 'S', 1762-1763, 'S', 1765-1766, 'S', 1768-1769, 'S', 1771-1772, 'S', 1774-1775, 'S', 1777-1778, 'S', 1780-1781, 'S', 1783-1784, 'S', 1786-1787, 'S', 1789-1790, 'S', 1792-1793, 'S', 1795-1796, 'S', 1798-1799, 'S', 1801-1802, 'S', 1804-1805, 'S', 1807-1808, 'S', 1810-1811, 'S', 1813-1814, 'S', 1816-1817, 'S', 1819-1820, 'S', 1822-1823, 'S', 1825-1826, 'S', 1828-1829, 'S', 1831-1832, 'S', 1834-1835, 'S', 1837-1838, 'S', 1840-1841, 'S', 1843-1844, 'S', 1846-1847, 'S', 1849-1850, 'S', 1852-1853, 'S', 1855-1856, 'S', 1858-1859, 'S', 1861-1862, 'S', 1864-1865, 'S', 1867-1868, 'S', 1870-1871, 'S', 1873-1874, 'S', 1876-1877, 'S', 1879-1880, 'S', 1882-1883, 'S', 1885-1886, 'S', 1888-1889, 'S', 1891-1892, 'S', 1894-1895, 'S', 1897-1898, 'S', 1900-1901, 'S', 1903-1904, 'S', 1906-1907, 'S', 1909-1910, 'S', 1912-1913, 'S', 1915-1916, 'S', 1918-1919, 'S', 1921-1922, 'S', 1924-1925, 'S', 1927-1928, 'S', 1930-1931, 'S', 1933-1934, 'S', 1936-1937, 'S', 1939-1940, 'S', 1942-1943, 'S', 1945-1946, 'S', 1948-1949, 'S', 1951-1952, 'S', 1954-1955, 'S', 1957-1958, 'S', 1960-1961, 'S', 1963-1964, 'S', 1966-1967, 'S', 1969-1970, 'S', 1972-1973, 'S', 1975-1976, 'S', 1978-1979, 'S', 1981-1982, 'S', 1984-1985, 'S', 1987-1988, 'S', 1990-1991, 'S', 1993-1994, 'S', 1996-1997, 'S', 1999-2000, 'S', 2002-2003, 'S', 2005-2006, 'S', 2008-2009, 'S', 2011-2012, 'S', 2014-2015, 'S', 2017-2018, 'S', 2020-2021, 'S', 2023-2024, 'S', 2026-2027, 'S', 2029-2030, 'S', 2032-2033, 'S', 2035-2036, 'S', 2038-2039, 'S', 2041-2042, 'S', 2044-2045, 'S', 2047-2048, 'S', 2050-2051, 'S', 2053-2054, 'S', 2056-2057, 'S', 2059-2060, 'S', 2062-2063, 'S', 2065-2066, 'S', 2068-2069, 'S', 2071-2072, 'S', 2074-2075, 'S', 2077-2078, 'S', 2080-2081, 'S', 2083-2084, 'S', 2086-2087, 'S', 2089-2090, 'S', 2092-2093, 'S', 2095-2096, 'S', 2098-2099, 'S', 2101-2102, 'S', 2104-2105, 'S', 2107-2108, 'S', 2110-2111, 'S', 2113-2114, 'S', 2116-2117, 'S', 2119-2120, 'S', 2122-2123, 'S', 2125-2126, 'S', 2128-2129, 'S', 2131-2132, 'S', 2134-2135, 'S', 2137-2138, 'S', 2140-2141, 'S', 2143-2144, 'S', 2146-2147, 'S', 2149-2150, 'S', 2152-2153, 'S', 2155-2156, 'S', 2158-2159, 'S', 2161-2162, 'S', 2164-2165, 'S', 2167-2168, 'S', 2170-2171, 'S', 2173-2174, 'S', 2176-2177, 'S', 2179-2180, 'S', 2182-2183, 'S', 2185-2186, 'S', 2188-2189, 'S', 2191-2192, 'S', 2194-2195, 'S', 2197-2198, 'S', 2200-2201, 'S', 2203-2204, 'S', 2206-2207, 'S', 2209-2210, 'S', 2212-2213, 'S', 2215-2216, 'S', 2218-2219, 'S', 2221-2222, 'S', 2224-2225, 'S', 2227-2228, 'S', 2230-2231, 'S', 2233-2234, 'S', 2236-2237, 'S', 2239-2240, 'S', 2242-2243, 'S', 2245-2246, 'S', 2248-2249, 'S', 2251-2252, 'S', 2254-2255, 'S', 2257-2258, 'S', 2260-2261, 'S', 2263-2264, 'S', 2266-2267, 'S', 2269-2270, 'S', 2272-2273, 'S', 2275-2276, 'S', 2278-2279, 'S', 2281-2282, 'S', 2284-2285, 'S', 2287-2288, 'S', 2290-2291, 'S', 2293-2294, 'S', 2296-2297, 'S', 2299-2300, 'S', 2302-2303, 'S', 2305-2306, 'S', 2308-2309, 'S', 2311-2312, 'S', 2314-2315, 'S', 2317-2318, 'S', 2320-2321, 'S', 2323-2324, 'S', 2326-2327, 'S', 2329-2330, 'S', 2332-2333, 'S', 2335-2336, 'S', 2338-2339, 'S', 2341-2342, 'S', 2344-2345, 'S', 2347-2348, 'S', 2350-2351, 'S', 2353-2354, 'S', 2356-2357, 'S', 2359-2360, 'S', 2362-2363, 'S', 2365-2366, 'S', 2368-2369, 'S', 2371-2372, 'S', 2374-2375, 'S', 2377-2378, 'S', 2380-2381, 'S', 2383-2384, 'S', 2386-2387, 'S', 2389-2390, 'S', 2392-2393, 'S', 2395-2396, 'S', 2398-2399, 'S', 2401-2402, 'S', 2404-2405, 'S', 2407-2408, 'S', 2410-2411, 'S', 2413-2414, 'S', 2416-2417, 'S', 2419-2420, 'S', 2422-2423, 'S', 2425-2426, 'S', 2428-2429, 'S', 2431-2432, 'S', 2434-2435, 'S', 2437-2438, 'S', 2440-2441, 'S', 2443-2444, 'S', 2446-2447, 'S', 2449-2450, 'S', 2452-2453, 'S', 2455-2456, 'S', 2458-2459, 'S', 2461-2462, 'S', 2464-2465, 'S', 2467-2468, 'S', 2470-2471, 'S', 2473-2474, 'S', 2476-2477, 'S', 2479-2480, 'S', 2482-2483, 'S', 2485-2486, 'S', 2488-2489, 'S', 2491-2492, 'S', 2494-2495, 'S', 2497-2498, 'S', 2500-2501, 'S', 2503-2504, 'S', 2506-2507, 'S', 2509-2510, 'S', 2512-2513, 'S', 2515-2516, 'S', 2518-2519, 'S', 2521-2522, 'S', 2524-2525, 'S', 2527-2528, 'S', 2530-2531, 'S', 2533-2534, 'S', 2536-2537, 'S', 2539-2540, 'S', 2542-2543, 'S', 2545-2546, 'S', 2548-2549, 'S', 2551-2552, 'S', 2554-2555, 'S', 2557-2558, 'S', 2560-2561, 'S', 2563-2564, 'S', 2566-2567, 'S', 2569-2570, 'S', 2572-2573, 'S', 2575-2576, 'S', 2578-2579, 'S', 2581-2582, 'S', 2584-2585, 'S', 2587-2588, 'S', 2590-2591, 'S', 2593-2594, 'S', 2596-2597, 'S', 2599-2600, 'S', 2602-2603, 'S', 2605-2606, 'S', 2608-2609, 'S', 2611-2612, 'S', 2614-2615, 'S', 2617-2618, 'S', 2620-2621, 'S', 2623-2624, 'S', 2626-2627, 'S', 2629-2630, 'S', 2632-2633, 'S', 2635-2636, 'S', 2638-2639, 'S', 2641-2642, 'S', 2644-2645, 'S', 2647-2648, 'S', 2650-2651, 'S', 2653-2654, 'S', 2656-2657, 'S', 2659-2660, 'S', 2662-2663, 'S', 2665-2666, 'S', 2668-2669, 'S', 2671-2672, 'S', 2674-2675, 'S', 2677-2678, 'S', 2680-2681, 'S', 2683-2684, 'S', 2686-2687, 'S', 2689-2690, 'S', 2692-2693, 'S', 2695-2696, 'S', 2698-2699, 'S', 2701-2702, 'S', 2704-2705, 'S', 2707-2708, 'S', 2710-2711, 'S', 2713-2714, 'S', 2716-2717, 'S', 2719-2720, 'S', 2722-2723, 'S', 2725-2726, 'S', 2728-2729, 'S', 2731-2732, 'S', 2734-2735, 'S', 2737-2738, 'S', 2740-2741, 'S', 2743-2744, 'S', 2746-2747, 'S', 2749-2750, 'S', 2752-2753, 'S', 2755-2756, 'S', 2758-2759, 'S', 2761-2762, 'S', 2764-2765, 'S', 2767-2768, 'S', 2770-2771, 'S', 2773-2774, 'S', 2776-2777, 'S', 2779-2780, 'S', 2782-2783, 'S', 2785-2786, 'S', 2788-2789, 'S', 2791-2792, 'S', 2794-2795, 'S', 2797-2798, 'S', 2800-2801, 'S', 2803-2804, 'S', 2806-2807, 'S', 2809-2810, 'S', 2812-2813, 'S', 2815-2816, 'S', 2818-2819, 'S', 2821-2822, 'S', 2824-2825, 'S', 2827-2828, 'S', 2830-2831, 'S', 2833-2834, 'S', 2836-2837, 'S', 2839-2840, 'S', 2842-2843, 'S', 2845-2846, 'S', 2848-2849, 'S', 2851-2852, 'S', 2854-2855, 'S', 2857-2858, 'S', 2860-2861, 'S', 2863-2864, 'S', 2866-2867, 'S', 2869-2870, 'S', 2872-2873, 'S', 2875-2876, 'S', 2878-2879, 'S', 2881-2882, 'S', 2884-2885, 'S', 2887-2888, 'S', 2890-2891, 'S', 2893-2894, 'S', 2896-2897, 'S', 2899-2900, 'S', 2902-2903, 'S', 2905-2906, 'S', 2908-2909, 'S', 2911-2912, 'S', 2914-2915, 'S', 2917-2918, 'S', 2920-2921, 'S', 2923-2924, 'S', 2926-2927, 'S', 2929-2930, 'S', 2932-2933, 'S', 2935-2936, 'S', 2938-2939, 'S', 2941-2942, 'S', 2944-2945, 'S', 2947-2948, 'S', 2950-2951, 'S', 2953-2954, 'S', 2956-2957, 'S', 2959-2960, 'S', 2962-2963, 'S', 2965-2966, 'S', 2968-2969, 'S', 2971-2972, 'S', 2974-2975, 'S', 2977-2978, 'S', 2980-2981, 'S', 2983-2984, 'S', 2986-2987, 'S', 2989-2990, 'S', 2992-2993, 'S', 2995-2996, 'S', 2998-2999, 'S', 3001-3002, 'S', 3004-3005, 'S', 3007-3008, 'S', 3010-3011, 'S', 3013-3014, 'S', 3016-3017, 'S', 3019-3020, 'S', 3022-3023, 'S', 3025-3026, 'S', 3028-3029, 'S', 3031-3032, 'S', 3034-3035, 'S', 3037-3038, 'S', 3040-3041, 'S', 3043-3044, 'S', 3046-3047, 'S', 3049-3050, 'S', 3052-3053, 'S', 3055-3056, 'S', 3058-3059, 'S', 3061-3062, 'S', 3064-3065, 'S', 3067-3068, 'S', 3070-3071, 'S', 3073-3074, 'S', 3076-3077, 'S', 3079-3080, 'S', 3082-3083, 'S', 3085-3086, 'S', 3088-3089, 'S', 3091-3092, 'S', 3094-3095, 'S', 3097-3098, 'S', 3100-3101, 'S', 3103-3104, 'S', 3106-3107, 'S', 3109-3110, 'S', 3112-3113, 'S', 3115-3116, 'S', 3118-3119, 'S', 3121-3122, 'S', 3124-3125, 'S', 3127-3128, 'S', 3130-3131, 'S', 3133-3134, 'S', 3136-3137, 'S', 3139-3140, 'S', 3142-3143, 'S', 3145-3146, 'S', 3148-3149, 'S', 3151-3152, 'S', 3154-3155, 'S', 3157-3158, 'S', 3160-3161, 'S', 3163-3164, 'S', 3166-3167, 'S', 3169-3170, 'S', 3172-3173, 'S', 3175-3176, 'S', 3178-3179, 'S', 3181-3182, 'S', 3184-3185, 'S', 3187-3188, 'S', 3190-3191, 'S', 3193-3194, 'S', 3196-3197, 'S', 3199-3200, 'S', 3202-3203, 'S', 3205-3206, 'S', 3208-3209, 'S', 3211-3212, 'S', 3214-3215, 'S', 3217-3218, 'S', 3220-3221, 'S', 3223-3224, 'S', 3226-3227, 'S', 3229-3230, 'S', 3232-3233, 'S', 3235-3236, 'S', 3238-3239, 'S', 3241-3242, 'S', 3244-3245, 'S', 3247-3248, 'S', 3250-3251, 'S', 3253-3254, 'S', 3256-3257, 'S', 3259-3260, 'S', 3262-3263, 'S', 3265-3266, 'S', 3268-3269, 'S', 3271-3272, 'S', 3274-3275, 'S', 3277-3278, 'S', 3280-3281, 'S', 3283-3284, 'S', 3286-3287, 'S', 3289-3290, 'S', 3292-3293, 'S', 3295-3296, 'S', 3298-3299, 'S', 3301-3302, 'S', 3304-3305, 'S', 3307-3308, 'S', 3310-3311, 'S', 3313-3314, 'S', 3316-3317, 'S', 3319-3320, 'S', 3322-3323, 'S', 3325-3326, 'S', 3328-3329, 'S', 3331-3332, 'S', 3334-3335, 'S', 3337-3338, 'S', 3340-3341, 'S', 3343-3344, 'S', 3346-3347, 'S', 3349-3350, 'S', 3352-3353, 'S', 3355-3356, 'S', 3358-3359, 'S', 3361-3362, 'S', 3364-3365, 'S', 3367-3368, 'S', 3370-3371, 'S', 3373-3374, 'S', 3376-3377, 'S', 3379-3380, 'S', 3382-3383, 'S', 3385-3386, 'S', 3388-3389, 'S', 3391-3392, 'S', 3394-3395, 'S', 3397-3398, 'S', 3400-3401, 'S', 3403-3404, 'S', 3406-3407, 'S', 3409-3410, 'S', 3412-3413, 'S', 3415-3416, 'S', 3418-3419, 'S', 3421-3422, 'S', 3424-3425, 'S', 3427-3428, 'S', 3430-3431, 'S', 3433-3434, 'S', 3436-3437, 'S', 3439-3440, 'S', 3442-3443, 'S', 3445-3446, 'S', 3448-3449, 'S', 3451-3452, 'S', 3454-3455, 'S', 3457-3458, 'S', 3460-3461, 'S', 3463-3464, 'S', 3466-3467, 'S', 3469-3470, 'S', 3472-3473, 'S', 3475-3476, 'S', 3478-3479, 'S', 3481-3482, 'S', 3484-3485, 'S', 3487-3488, 'S', 3490-3491, 'S', 3493-3494, 'S', 3496-3497, 'S', 3499-3500, 'S', 3502-3503, 'S', 3505-3506, 'S', 3508-3509, 'S', 3511-3512, 'S', 3514-3515, 'S', 3517-3518, 'S', 3520-3521, 'S', 3523-3524, 'S', 3526-3527, 'S', 3529-3530, 'S', 3532-3533, 'S', 3535-3536, 'S', 3538-3539, 'S', 3541-3542, 'S', 3544-3545, 'S', 3547-3548, 'S', 3550-3551, 'S', 3553-3554, 'S', 3556-3557, 'S', 3559-3560, 'S', 3562-3563, 'S', 3565-3566, 'S', 3568-3569, 'S', 3571-3572, 'S', 3574-3575, 'S', 3577-3578, 'S', 3580-3581, 'S', 3583-3584, 'S', 3586-3587, 'S', 3589-3590, 'S', 3592-3593, 'S', 3595-3596, 'S', 3598-3599, 'S', 3601-3602, 'S', 3604-3605, 'S', 3607-3608, 'S', 3610-3611, 'S', 3613-36